BLS WORKING PAPERS

U.S. Department of Labor
U.S. Bureau of Labor Statistics
Office of Employment Research and Program Development

Outsourcing, Occupationally Homogeneous Employers, and Growing Wage Inequality in the United States

Elizabeth Weber Handwerker, U.S. Bureau of Labor Statistics

Working Paper 522
March 2020

JEL codes: J31, J53, M51, L24, M55
Keywords: outsourcing, employer homogeneity, wage inequality, contracting out, employment polarization

Outsourcing, Occupationally Homogeneous Employers, and Growing Wage Inequality in the United States

Elizabeth Weber Handwerker
U.S. Bureau of Labor Statistics

March, 2020

Abstract: This paper develops measures of the occupational homogeneity of employers as indicators of outsourcing. Findings are threefold. First, workers in low-wage occupations saw their employing establishments become more occupationally homogeneous during 2002-2016. Second, wages are strongly related to occupational homogeneity, particularly for workers in low-wage occupations. Third, changes in the occupational homogeneity of workplaces can explain nearly half of increased wage inequality among workers in the lower 98% of the wage distribution over this period. The growing separation of workers in low-wage occupations into different employers from workers in high-wage occupations is an important part of wage inequality growth.

I am grateful to seminar participants at the American Economic Association 2014 Annual Meetings, the University of Maryland, the Economic Policy Institute, the Federal Reserve Board of Governors, the Society of Labor Economists 2016 and 2017 Annual Meetings, the Center for Equitable Growth, the NBER 2018 Summer Institute, the BRIQ Institute, and the Bureau of Labor Statistics for comments, with special thanks to Matt Dey, Katharine Abraham, David Autor, Erica Groshen, Sue Houseman, David Levine, Anne Polivka, Alex Bryson, David Piccone, David Weil, Nicole Fortin, Fernando Rios-Avila, and anonymous referees for particularly helpful comments and suggestions. This paper grew out of a project begun jointly with James Spletzer, and I thank him for his many contributions. Research assistance was provided by Lowell Mason. Any opinions and conclusions herein are those of the author and do not necessarily reflect the views or policies of the U.S. Bureau of Labor Statistics.

I. Introduction

Growing inequality of wages, particularly between employers, has been a key feature of the labor market in recent decades. Many changes in the labor market have been examined as potential sources of this inequality growth—including the decline of manufacturing, the role of technology in replacing employer demand for routine work, and the increased potential for imported goods and services to replace domestic production. This paper examines an additional source of growing wage inequality: the changing distribution of occupations between employers as the organization of production changes, with employers retaining certain types of work within the workplace and outsourcing other work.

Much evidence shows that establishments play an important role in determining individual wages, beyond the role of individual workers' characteristics (Groshen 1991a, 1991b; Bronars and Famulari 1997; Abowd, Kramarz, and Margolis 1999; Lane, Salmon, and Spletzer 2007; Card, Heining, and Kline 2013). Several authors have used employer microdata to study growing variability in earnings in the U.S. from the mid-1970s to the early 2000s, and have found it due more to variation between establishments than to variation within establishments (Davis and Haltiwanger 1991; Dunne, Foster, Haltiwanger, and Troske 2004; Barth, Bryson, Davis, and Freeman 2016; Handwerker and Spletzer 2016; and Song, Price, Guvenen, Bloom, and von Wachter 2016),[1] while the increased sorting of high-paid workers to high-paying employers drives much of the growth in pay inequality between employers (Song, Price, Guvenen, Bloom, and von Wachter, 2016). The results in this paper show that occupational homogeneity—a specific form of worker sorting—is a key explanation for this growth in between employer wage inequality. More and more workers in low-wage occupations are employed in different workplaces from workers in other occupations, exacerbating differences in their pay.

The intersection of growing underlying wage inequality and the business environment in the United States can make it profitable for employers to focus on employing either low or high wage workers. Growing wage inequality among workers has arisen from such sources as the changing composition of the workforce and changing returns to education and experience,[2] the growing inequality within education and skill groups[3], and the differential impact of technology on the worker skill distribution[4]. As wages for different kinds of work become less equal, employers face regulations requiring nondiscrimination across employees in the coverage of pension, health insurance and other benefits (EBRI 2009, Perun 2010),[5] increasing incentives to contract out work that pays very different wages from the work of other employees. Moreover, social norms may make it more acceptable for employers to contract out work rather than pay very different wages to employees doing different kinds of work (Weil 2014).

[1] There is a large and growing literature on wage inequality growth in Europe, based on employee-employer linked data, including Card, Heining, and Kline (2013), who emphasize the role of increased worker sorting between employers in explaining wage inequality growth in Germany.
[2] Bound and Johnson 1992, Katz and Murphy 1992, Lemieux 2006
[3] Juhn, Murphy, and Pierce 1993, Katz and Autor 1999
[4] Juhn, Murphy, and Pierce 1993, Acemoglu 2002, Autor, Katz, and Kearney 2006, 2008
[5] Perun (2010) lists a variety of employment benefits which receive favorable tax treatment and are required to be available to low-wage as well as high-wage employees of each employer.

Other potential reasons for businesses to outsource work include increasing ability to smooth workload, economies of scale available to providers of specialized services (Abraham and Taylor, 1996), or a focus on "core competencies" enabled by technologies for specifying and monitoring work done by outsiders (Weil 2014). However, to the extent that labor cost savings and avoidance of efficiency wages or rents for occupations with low wages in the labor market are key reasons for outsourcing, it can lead to employers specializing in high or low-wage work, and result in growing wage inequality between establishments. Goldschmidt and Schmeider (2015) show labor cost savings to be a primary reason for outsourcing in Germany, as outsourced workers lose firm-specific rents. In three well-defined occupational categories, they find that losses of these firm-specific rents can account for 9% of all growth in German wage inequality from 1985 to 2008.

In U.S. data, direct measures of outsourcing are not generally available. Researchers have instead focused on particular industries or occupations associated with performing support tasks for other businesses. Dey, Houseman, and Polivka (2010) show a marked increase in various measures of outsourcing in recent years such as trends in temporary help or employment services. Estimates from several sources show these industries roughly doubling in size from 1992 to 2002. They also document an increase in the employment share of occupations associated with outsourced labor services, such as school bus and truck drivers in the transportation industry and accountants in the business services industry. Yet these measures only capture a fraction of outsourcing—that which occurs in these specific industries. Dube and Kaplan (2010) use individual-level data to show the impact of outsourcing on wages and benefits for janitors and guards, but again, their measures can only capture outsourcing of a narrow set of occupations.

This paper develops economy-wide measures of outsourcing, using the homogeneity of occupations by employer, as measured in the detailed microdata of the Occupational Employment Statistics Survey conducted by the Bureau of Labor Statistics. These measures distinguish between two types of outsourcing, which may have differing impacts on wage inequality. When businesses outsource work to avoid monitoring, hiring, or other costs for occupations in which they have less expertise, there will be less variety overall in the occupations they employ. However, when businesses outsource work to narrow the wage distribution of their employees, the variance of wages predicted from the particular set of occupations they employ will decrease. The impact of the changing distributions of occupations and of employer occupational homogeneity are compared with the effects of other changes in employer characteristics (industry, size, and location) on the overall distribution of wages.

There are three major findings. First, from November 2002 through November 2016, the occupational homogeneity of employers increased for workers in typically low-wage occupations, after controlling for other employer characteristics. Second, wages are related to the occupational homogeneity of establishments. Workers in more occupationally homogeneous establishments earn lower wages. This relationship holds even after controlling for workers' own occupations and observable characteristics of their employers, and is strongest for workers in occupations typically paid low wages. Third, changes in the distribution of this occupational homogeneity are related to the growth in private-sector wage inequality observed in the data during this time period. Nearly half of the growth in ln(wage) variance, as measured in the OES data, can be attributed to the growing occupational homogeneity of establishments over this period, and the growing occupational homogeneity of workplaces also plays an important role in explaining the growth of wage inequality between employers. A measure of employer homogeneity that focuses on the distribution of occupations by wage levels matters more for the growth in wage inequality than a more functional measure of employer homogeneity that ignores wage differences among occupations.

The paper is organized as follows: Section II describes measures of occupational homogeneity. Section III describes trends in measured occupational homogeneity of employers. Section IV describes relationships between employer occupational homogeneity and employee wages. Section V describes the impact of the changing distributions of occupation and the occupational homogeneity of employers on wage inequality over time. Section VI concludes.

II. Measuring the Occupational Homogeneity of Employers

I use the term "occupational homogeneity"[6] to describe the variety of occupations employed at a place of business, separate from the tasks performed by individual employees (their occupations), the type of work done at the business (its industry) or the size of the business. Much scholarship on outsourcing (for example Dey, Houseman, and Polivka, 2010; and Erickcek, Houseman, and Kalleberg, 2003) examines particular occupations and particular industries. In contrast, occupational homogeneity is intended as a measure of the variation in work done in all businesses, through the full range of industries in the economy. This section defines two measures of occupational homogeneity and presents evidence showing that these measures are related to examples in the outsourcing literature.

The two measures of the occupational homogeneity of establishments are very different: (1) a measure involving the overall distribution of occupations, regardless of whether they are high or low paid, and (2) a measure that explicitly models the variation in wages of establishments due to the distribution of occupations employed.

The first measure of occupational homogeneity for establishment *j* at time *t* is constructed with a Herfindahl-Hirschman index of employment, *n*, in each occupation *k* within that establishment:

$$(1) \qquad H_{jt} = \sum_{k=1}^{100} \left(\frac{n_{kjt}}{n_{jt}}\right)^2$$

This index uses the 100 minor occupational categories at the 3-digit level of the Standard Occupational Classification system.[7] It varies from 1/100 (equal representation of all occupations) to 1 (complete homogeneity). Increased occupational homogeneity at the establishment level by this measure indicates that employers are becoming more specialized,

[6] Earlier versions of this paper referred to the same concept as "occupational concentration."

[7] Handwerker and Spletzer (2016) studied this type of general occupational homogeneity with Herfindahl-Hirschman indices, using both the detailed 6-digit occupations of the Standard Occupational Classification System (829 categories) and the 2-digit major occupational categories of the Standard Occupational Classification System (22 categories), and found very similar time trends and relationships between occupational classification and wages with broad and detailed versions of this measure.

consistent with outsourcing work to other employers. Trends in this measure indicate whether establishments throughout the U.S. economy are becoming more homogeneous in the occupations they employ. However, this measure cannot distinguish between specializing in a few occupations typically paid very different wages, such as 29-1000 (Healthcare Diagnosing or Treating Practitioners) and 31-1100 (Home Health and Personal Care Aides; and Nursing Assistants, Orderlies, and Psychiatric Aides), or specializing in a similar number of occupations that are typically paid very similar wages.

In contrast, the second measure of occupational homogeneity is explicitly constructed to capture the similarity or dissimilarity of typical wages for the occupations each establishment employs. It is the variance of wages for each establishment that would be predicted from the establishment's distribution of employment by occupation, without using information on the actual wages paid at the establishment. Using average log wages for each minor occupational category in each time period, the log wage paid by employer *j* to worker *i* in occupation *k* at time *t* is estimated as $\hat{w}_{ijt} = \bar{w}_{kt} + \varepsilon_{ijt}$, where $\bar{w}_{kt}$ is the mean log wage for all employees in occupation *k* at time *t* and ε_{ijt} is distributed normally, with mean 0 and standard deviation σ_k. From the occupational distribution of employer j at time t, the estimated mean log wage for j at t is estimated $\hat{\bar{w}}_{jt} = \frac{\sum_k \sum_{i \in k} \bar{w}_{kt}}{n_{jt}}$, where n_{jt} is the total employment for employer *j* at time *t*, and $i \in k$ denotes observations in which individual *i* has occupation *k*. Again, using only the distribution of occupations employed and the average wages of these particular occupations across all employers at time t, the predicted log wage variance for employer *j* at time *t* is

$$(2) \qquad \hat{V}_{jt} = \frac{\sum_i (\hat{w}_{ijt} - \hat{\bar{w}}_{jt})^2}{n_{jt}} = \frac{\sum_k n_{jkt} [(\bar{w}_{kt} - \hat{\bar{w}}_{jt})^2]}{n_{jt}} + \frac{\sum_k n_{jkt} \sigma_{kt}^2}{n_{jt}}.$$

This has two terms: the variation in average wages between occupations, and the average of within-occupation log wage variances. The sum of these terms is the second measure of occupational homogeneity: the variance of log wages for employer *j* at time *t* predicted from the composition of the occupations employed.

Both of these measures are estimated with the microdata of the Occupational Employment Statistics (OES) Survey for the private sector in the United States for 2002 to 2016.[8] These microdata record the number of employees by wage interval within detailed occupation categories for hundreds of thousands of establishments per year. The OES survey is designed to produce estimates of employment and wages in the United States for each detailed occupation, by geography and industry. It covers all establishments in the United States except for those in agriculture, private households, and unincorporated self-employed workers without employees. It is the only survey of its size and scope.

The OES collects data for a sample of about 200,000 establishments each November and each May. Sampled establishments are asked to report the number of employees in each occupation by wage interval. As described in Dey and Handwerker (2016), the OES uses a complex sample design intended to minimize the variance of published wage estimates for each occupation within industries and geographic areas. Establishments expected to employ rarer

[8] An earlier version of this paper used microdata for 1999 to 2015. However, as described by Abraham and Spletzer (2010), many first-line supervision occupations in establishments of less than 500 workers were erroneously coded as managerial occupations during 1999-2001. Thus, microdata for earlier years have now been dropped.

occupations or occupations with greater variation in wages have relatively larger probabilities of selection.

In using OES data to study wage inequality, it is important to understand that the OES data *cannot* measure inequality in the topmost percentiles of the wage distribution. Wages are reported to the OES in intervals. The OES program uses the mean of each wage interval every year from the National Compensation Survey (NCS) to assign wages for employees in each interval. Earnings of individuals at the very top of the wage distribution are topcoded in the OES—the uppermost interval in the recent OES surveys is "$208,000 and over." Averaged across all years, the uppermost interval contains roughly 1.3 percent of employment. Handwerker and Spletzer (2014) compare wage inequality levels and trends in these same reweighted OES microdata with the wage inequality level and trends in the outgoing rotation group microdata of the CPS, which has been used in many of the most cited studies of wage inequality. They show the interval nature of wage collection in the OES has almost no impact on overall wage variance trends. The reweighted OES data broadly replicate the CPS wage distribution trends: overall wage variances in each year are similar in the reweighted OES and CPS microdata, as well as overall variance trends, and variance trends by sector, industry groups, and occupation groups.

The OES sample design uses 3 years, or 6 panels of data collection, to produce detailed published estimates of employment and wages. It is not designed to produce time series estimates of either employment or wages for any individual occupation. This paper uses data reweighted using the methodology of Abraham and Spletzer (2010) to match the detailed industry and employer size distribution of the Quarterly Census of Employment and Wages for each May and November.

Establishments are the sampling units of the OES, and so this paper focuses on measures of occupational homogeneity at the establishment level. However, all the main results in this paper have been repeated with measures constructed at the Employer Tax-ID level (EIN), with little impact on the results (EIN-level results are shown in Appendix C).

The Data Appendix contains summary statistics, including the composition of occupations and industries. The average worker has an inflation-adjusted wage of $16.20/hour, or a ln(wage) of 2.56, and is observed in an establishment with a measured ln(wage) variance of 0.154. The average Hirfindahl-Hirschman index for workers' establishments is 0.408, and the average predicted variance of ln(wages) estimated from its workers' occupational composition is 0.270. It is unsurprising that the predicted ln(wage) variance based only on the occupations employed at the establishment is higher than the measured ln(wage) variance because of the large literature describing the impact of employer-specific factors on wages.

Table 1 compares the two measures of establishment-level occupational homogeneity for several occupation-industry groups studied as examples of outsourcing by Abraham and Taylor (1996); Dube and Kaplan (2010); Dey, Houseman, and Polivka (2010); Weil (2014); and Goldschmidt and Schmeider (2015): the entire food preparation and serving major occupational group, janitors, security guards, truck drivers, accountants, computer occupations, engineering occupations, and lawyers. Outsourcing of workers in these occupations means that they are employed in the specialty industries of food services, janitorial services, security guard services, truck transportation, accounting services, computer services, engineering services, or law offices, rather than the industry of the business to which they provide these services. Table 1 shows that for every single one of these example occupations or occupation groups, the Herfindahl-Hirschman indices for employers of these workers (as defined in equation (1)) are higher, on average, indicating greater occupational homogeneity of employers, when they are employed in their specialty industry than when they are employed in other industries. Moreover, for every example occupation except lawyers (the smallest, highest paid, and employed by the most occupationally homogeneous employers in its specialty industry), the predicted variances of wages based on the occupational distribution of their employers are lower, on average, indicating greater occupational homogeneity of employers, when they are employed in their specialty industry than when they are employed in other industries. Both measures of occupational homogeneity measures defined in this section—designed to measure outsourcing across all occupations and industries—indicate greater occupational homogeneity in the relevant industries to which workers are outsourced for the specific occupations studied in the outsourcing case-study literature.

III: Trends in Occupational Homogeneity Measures

Understanding trends in occupational homogeneity measures is complicated by contemporaneous changes in the overall occupational composition of the labor force. As described by Autor, Katz, and Kearney (2006, 2008), among others, employment in typically low-wage and typically high-wage occupations has increased, while employment in many typically middle wage-occupations has decreased. Figure 1 shows employment over time for occupational quintiles in the OES. Employment polarization is clear in the OES data: there is an increasing fraction of employment over time in the top and bottom quintiles, with a decreasing fraction of employment in the middle three quintiles. This polarization means that if we entirely ignore the grouping of employment into establishments, the portion of the variance of ln(wages) for all workers due to wage variation between occupations is generally increasing—from values varying between 0.20 and 0.21 in the early years of the microdata to values varying between 0.22 and 0.23 in later years. In other words, the overall polarization of employment mechanically leads to increases in the variance of wages between occupations. There is no such mechanical relationship between overall changes in employment by occupation and the Herfindahl-Hirschman index: a version of the Herfindahl-Hirschman index that pools workers across all employers varies only between .027 and .028 over this period, with no clear time trend.

The actual time trend of mean occupational homogeneity at the establishment level is described with regressions of the form

$$(3)\qquad OccHomogeneity_{ijt} = \alpha Survey\ Date_t + \beta I(May\ Survey_t) + \gamma X_{ijt} + \varepsilon_{ijt}$$

where SurveyDate measures time in decades since November 2002, I(May Survey) is an indicator to capture seasonal variation between data collection in May and November, and X_{ijt} are other observable characteristics of individual *i* (occupation) and employer *j* (industry, geography, and size) at time *t*. Trend regression results for equation (3) are shown in Tables 2 and 3. The first two rows of Table 2 show an increase over time in the Herfindahl-Hirschman measure of the occupational homogeneity of employers overall, but changes in occupations and employer characteristics explain about 95% of this increase. Meanwhile, the predicted variance of ln(wages) measure of occupational homogeneity has risen over time, showing a trend of decreasing employer homogeneity overall for this measure.

Further rows of Table 2 repeat this analysis for subgroups of occupations. Occupations (at the 3-digit minor occupational category SOC level) are grouped by average wage into quintiles, with roughly equal total weighted employment in each quintile. [9] Appendix A lists the occupations of each quintile, while counts of the individual and employer observations for each quintile are in the Data Appendix. The list of occupations in the lowest-paid quintile is a short one, because the occupations in this quintile tend to be large, such as Food and Beverage Serving Workers. The list of occupations in the highest-paid quintile is much longer, because these occupations tend to be smaller, such as Social Scientists.

The subgroup rows of Table 2 show the greatest increases over time in the Herfindahl-Hirschman measure of occupational homogeneity—after including controls for occupation and establishment characteristics—occur in the bottom and top quintiles of occupations. For the predicted variance measure of occupational homogeneity, there is an increase over time for the top three higher-paid quintiles of occupations that persists after including the controls, and a decrease in variance over time (showing greater employer homogeneity) that persists after including the controls for the occupations in the lowest-paid quintile.

The trend regressions of Table 2 measure changes only in the mean levels of these occupational homogeneity measures. In overall distributions of the two measures of occupational homogeneity, the Herfindahl-Hirschman measure has a similar shape over time. However, the distribution of values for the predicted variance of ln(wages) becomes increasingly bimodal over time, with modal values falling for establishments employing people in typically low-wage occupations and rising for employers of higher-wage occupations.

Figure 2 uses the same five quintiles of occupations by typical wages used in Table 2, and shows the fraction of workers in each quintile of occupations who work in establishments without any workers in other quintiles. It is unsurprising that workers in all other quintiles of occupation are growing less likely to have any coworkers in the middle three quintiles, as the middle quintile occupations have declining shares of overall employment over time. However, Figure 2 shows that workers in the bottom three quintiles increasingly have no coworkers in the top quintile of occupation, and workers in the top three quintiles increasingly have no coworkers in the bottom quintile of occupation, although the bottom and top quintiles of the occupational distribution have increasing shares of overall employment over time.

This figure helps to explain how the predicted-variance measures can be falling (increased homogeneity) for workers in typically low-wage occupations while the same measures can be rising (increased heterogeneity) for workers in typically high-wage occupations. The polarization of overall employment is increasing, with rising shares of employment in the

[9] To form quintiles, occupations are ranked by their average wages across all years. This grouping of occupations is quite stable over time.

top and bottom quintiles of occupations at the expense of the middle three quintiles of occupations. However, this polarization is not happening evenly across establishments.

To illustrate the impact of these trends in employment by occupational quintiles on the predicted variance of wages for establishments, consider coarsening the occupational distribution into only three occupation groups: low-wage occupation group L, middle-wage occupation group M, and high-wage occupations H, with mean wages for occupations in each group $\bar{\bar{w}}_L < \bar{\bar{w}}_M < \bar{\bar{w}}_H$ and within-occupations wage variances by group $\sigma_L^2 < \sigma_M^2 < \sigma_H^2$. Each establishment j contains $n_L \geq 0$ workers in the low-wage occupation group, $n_M \geq 0$ workers in the middle-wage occupation group, and $n_H \geq 0$ workers in the high-wage occupation group, with $n_L + n_M + n_H = n_j$. The predicted variance of wages for each establishment is $\hat{V}_j = \frac{n_L[(\bar{w}_L - \hat{y}_j)^2]}{n_j} + \frac{n_M[(\bar{w}_M - \hat{y}_j)^2]}{n_j} + \frac{n_H[(\bar{w}_H - \hat{y}_j)^2]}{n_j} + \frac{n_L\sigma_L^2}{n_j} + \frac{n_M\sigma_M^2}{n_j} + \frac{n_H\sigma_H^2}{n_j}$.

For workers in occupation group L, employing establishments have higher n_L, lower n_M, and lower n_H, and, as shown in Figure 2, growing numbers of workers in occupation group L work in establishments with $n_M = n_H = 0$. There is little variation in wages between the low-wage occupations (a low value of σ_L^2), which reduces the typical values of $(\bar{w} - \hat{y})$. With fewer workers in middle or high wage occupations, there is less weight on the other components of the predicted wage variance. This lowers $\hat{V}$ for the workers in occupation group L.

For workers in occupation group H, employing establishments have lower n_L, lower n_M, and higher n_H, and, as shown in Figure 2, growing numbers of workers in occupation L work in establishments with $n_L = n_M = 0$. Although average wages are higher in these establishments, reducing the typical values of $(\bar{w} - \hat{y})$, the greater weight n_H associated with the high wage variance within these occupations, σ_H^2, means a greater $\hat{V}$ overall for these establishments.

To understand how a high-paying occupation can have growing occupational homogeneity over time by the Herfindahl-Hirschman measure at the same time it has shrinking occupational homogeneity over time by the predicted wage variance measures (increased variance), consider accountants, an example high-wage occupation from Table 1. During this period, the share of accountants working in the manufacturing sector fell while the share working in the management of companies sector rose. Establishments that employed accountants became increasingly specialized in two occupational categories: Health Diagnosing and Treating Practitioners and Business Operations Specialists, with some growth in the number of Financial Specialists, the category that includes accountants. All of these are high-wage occupational categories. Meanwhile, employers of accountants employed fewer people in occupational categories such as Other Office and Administrative Support Workers, Production occupations, Secretaries and Administrative assistants, and Financial Clerks. This changing occupational distribution for employers of accountants meant rising predicted wage variances (more heterogeneity by wage variance measures), even as these employers concentrated employment in particular occupations (greater homogeneity by the Herfindahl-Hirschman measure).

Weil (2014) describes how large corporations have shed many low-wage tasks by outsourcing them to other companies, which repeatedly subcontract them to smaller and smaller employers. Figure 3 shows that establishment size plays a role in the increasing segregation of workers in the lowest-paid quintile of occupations and workers in the highest-paid quintile of occupations into separate establishments, following the pattern Weil describes. Rising shares of employment for the lowest-paid quintile of occupations occurred only in establishments of less than 100 workers, while rising share of employment for the highest-paid quintile of occupations occurred more sharply in establishments of 100 or more workers.[10]

Table 3 shows the implication of this growing segregation of workers by establishment size for time trends in measured establishment occupational homogeneity. This table disaggregates the results of equation (3) by both establishment size and occupation group. For workers in the lowest-paid quintile of occupations in establishments with less than 100 workers, including the controls described above, workplaces are increasingly homogenous, by both measures. Furthermore, the predicted variance of ln(wages) measure of occupational homogeneity has fallen faster (homogeneity has increased more) for low-paid workers in these smaller establishments than in establishments with 100 or more workers.

Appendix B describes employer occupational homogeneity trends along several other dimensions—by state-level unionization rates, more detailed establishment size categories, establishment age, industrial sector, and Employer Tax Identification Number (EIN) size.

For workers in low-wage occupations, all measures show a clear trend of increased employer occupational homogeneity over time. The next section shows these workers' wages are lower when they work for more occupationally homogenous employers.

IV: Relationships between Measured Occupational Homogeneity and Wages

The outsourcing literature provides several examples of occupations in which outsourcing is associated with lower wages, including occupations listed in Table 1. Among the example occupations in Table 1, all of the low wage occupations (food preparation and service, janitors, and security guards) earn considerably lower average wages in outsourced specialty industries than in other industries. These example occupations are examples precisely because there are obvious industries to which they can be outsourced; most other occupations do not have such obvious industries for outsourcing. However, the advantage of occupational homogeneity is that is can be measured for the employers of all occupations. This section shows the relationship between occupational homogeneity and wages for all workers.

I describe the relationship between occupational homogeneity and wage with regressions of the form

(4) $$\ln(wage_{ijt}) = \alpha OccHomogeneity_{jt} + \beta Date_t OccHomogeneity_{jt} + \gamma X_{ijt} + \varepsilon_{ijt},$$

[10] Patterns are similar for establishments of 1-49 workers and establishments of 50-99 workers. Patterns are also quite similar when using EIN size instead of establishment size.

where $OccHomogeneity_{jt}$ is the measure of occupational homogeneity for the employer of individual *i* at employer *j* in time t, $Date_t$ is the date, and X_{ijt} are other observable characteristics of individual i (occupation) and employer j (industry, geography, and size) at time t. Results of this regression are shown in Table 4. The first row of this table gives estimates of the impact of occupational homogeneity on wages, α, with no additional variables. These estimates show that increased occupational homogeneity is associated with lower wages overall. Estimates of the coefficients β (not shown) indicate that all these relationships have significantly strengthened over time. The second row of Table 4 gives these estimates with additional variables added to the regression. These detailed controls reduce the magnitude of the relationship between occupational homogeneity and wages, α, but the estimates maintain the same sign and remain very significant.

Further rows of Table 4 repeat this analysis for the same subgroups of occupations as in Table 2. The relationship between occupational homogeneity and wages, for both measures of occupational homogeneity, is generally stronger for workers in typically low-wage occupations than for workers in typically high-wage occupations. The relationship between the typical wage levels for a quintile of occupations and the wage coefficient of occupational homogeneity for the occupations in that quintile is not monotonic, with the largest wage coefficients are generally for the quintile of occupations with the second-lowest typical wages. For workers in the highest-paid quintile of occupations, greater occupational homogeneity is associated with *higher* wages by both measures of occupational homogeneity, once industry, own-occupation, geography and employer size are taken into account.

Estimates of β (not shown) for all occupations, with full occupation and employer controls, have the opposite sign to the estimates of α, indicating that the impact of occupational homogeneity on wages for all occupations has been decreasing over time overall. However, the regressions by quintile of occupation show that these estimates of β reverse for the middle-wage occupations, indicating that the relationship between occupational homogeneity and wages has been growing over time for these occupations.

Appendix B describes further heterogeneity in the relationship between occupational homogeneity and wages—by state-level unionization rates, establishment size, establishment age, industrial sector, and Employer Tax Identification Number (EIN) size.

This section has described the relationship observed between occupational homogeneity and wages, but does not say employer homogeneity "causes" lower wages for workers in lower-wage occupations. The data used in this paper do not allow me to measure whether differences in unmeasured skills and tasks—within the same occupation—might explain some of the difference in wages between workers in more and less homogenous workplaces. For example, janitors who work in the janitorial services industry may lack some specialized skills of janitors in other industries, and may perform somewhat different tasks than those employed in other industries. However, the many U.S. examples described in Weil (2014) and the labor force histories of German workers whose jobs are outsourced, as documented in Goldschmidt and Schmieder (2015) provide evidence that some portion of the observed relationship between employer homogeneity and wages is causal. The estimates in thus section should thus be considered an upper bound for the size of the causal impact of employer homogeneity on wages.

V. Occupational Homogeneity and Wage Inequality

The association between occupational homogeneity and lower wages for workers in lower-wage occupations, coupled with the trend of growing occupational homogeneity for workers in the lowest-wage occupations, suggests a role for occupational homogeneity in explaining growing wage inequality. Barth, Bryson, Davis, and Freeman (2016) highlighted that most inequality growth is between establishments, and is not explained by industry or geography. Moreover, Song, Price, Guvenen, and von Wachter (2016) show that the vast majority of pay-inequality growth at small and medium-sized firms from 1978-2013 was due to increasing segregation and sorting of workers who earn higher pay—without describing what about these workers makes them higher-paid workers—to firms that pay higher wages. Weil (2014) speculated that increased fissuring of employers could exacerbate wage inequality, but he did not have data to measure this directly. This section presents evidence showing that changes in occupational homogeneity can explain a substantial amount of the overall growth in wage inequality during this period.

I use Rios-Avila's implementation of Fortin, Firpo, and Lemiuex's Recentered Influence Function (RIF) Decomposition method to decompose changes in real ln(wage) variance from Nov 2002 & May 2003 to Nov 2014 & May 2015[11] into portions that can be explained by the changing composition of workers by occupation, and the changing composition of their employing establishments by industry, geography, size, and occupational homogeneity. Because the occupational homogeneity measures are continuous rather than categorical variables, these variables are divided into quartiles for this reweighting exercise.[12] The evidence in Table 2 shows that occupational homogeneity is changing in different ways for different quintiles of occupations. Thus, I interact occupational homogeneity variables with the same quintiles of occupation used above.[13] In addition, I add a dummy variable for lowest-wage quintile occupations employed establishment of less than 100 workers that are in the bottom half of the predicted variance distribution to the vector of indicator variables describing the predicted variance measure of occupational homogeneity.

Results are shown in Table 5. The changing composition of employment by industry, geography, establishment sizes, occupational quintiles, and the categories of occupational homogeneity described above can more than explain all of the change in ln(wage) variance from 2002-2003 to 2014-2015. The specification error is only 2% of this amount. Decomposing the change in ln(wage) variance by source, only tiny amounts are explained by changes in

[11] The November 2014 – May 2015 period is chosen as the end date because overall wage variance in lower in the 2016 panels than in the 2015 panels for the unimputed OES microdata. Results (without bootstrapped standard errors) for the later end date are in Appendix D.

[12] Quartiles are chosen for the main results here so that the interaction between occupational homogeneity quartiles and occupational quintiles is 20 categories, roughly the same number of categories as the number of industrial sectors. If I instead divide the measures of occupational homogeneity into deciles, and use the same detailed levels of industry and occupation variables as in Tables 2-4, the interaction between occupational homogeneity and occupational quintiles explains a similar amount of wage variance growth, but the decomposition takes so long to run that bootstrapping the standard errors is no longer possible. Results are shown in Appendix D.

[13] This follows the example of Goldschmidt and Schmieder (section V.C.), who use indicators for deciles of the firm wage effect interacted with dummies for frequently outsourced occupations.

employment by industry, geography, and establishment size. In contrast, 51% of the change in ln(wage) variance can be explained by changes in the occupational quintiles of workers (the polarization of employment). 6% can be explained by changes in the occupational homogeneity of employers by the Herfindahl-Hirschman measure, and 41% of the growth in ln(wage) variance is explained by changes in the occupational homogeneity of employers by the predicted variance of wages for the occupations they employ.

Dinardo, Fortin, and Lemiuex advise that their method should be implemented with bootstrapped standard errors. The results in Table 5 do not yet have these bootstrapped standard errors, but these are now being computed on the fastest computer available to me, and should finish late in the summer of 2020, in time to be incorporated into the next revision of this paper.

To examine the impact of changing occupational homogeneity on the growth of wage variance between establishments, I use the Dinardo-Fortin-Lemiuex (DFL) 1996 method. This method calculates counterfactual wage distributions by reweighting observable characteristics in the later period to their distributions in the earlier period. In this exercise, increased wage inequality growth from the November 2002 – May 2003 period to the November 2014 - May 2015 period can be attributed to changes in the distribution of employment by reweighting the data in the later period so that characteristics have the same distribution that they did in the earlier period. The DFL method cannot attribute changes in the wage distribution to the amounts attributable to changes in each characteristic as the FFL method does, but producing this counterfactual distribution of wages makes it simple to examine the impact of this change on the between-establishment wage variance.

The overall variance of real ln(wages) increased from 0.375 in 2002-2003 to 0.400 in 2014-2015, and essentially all of this increase is due to between-establishment wage variance increasing from 0.219 to 0.243. Reweighting the 2014-2015 data to the 2002-2003 distribution of employment by (1) the Herfindahl-Hirschman measure of occupational homogeneity interacted with quintiles of the occupational distribution, (2) the predicted variance of establishment ln(wages), and (3) an indicator for workers in typically low-wage occupations employed in small homogenous establishments can explain 55% of wage variance growth between establishments

In sum, these results show that changes in occupational homogeneity are a very important part of growing wage inequality for the lower 98.5% of the wage distribution. Furthermore, the predicted variance of wages measure matters more for wage inequality than Herfindahl-Hirschman measure. The separation of typically-low wage occupations into separate workplaces from typically-high wage occupations is the form of occupational homogeneity that matters for wage inequality.

VI. Summary: Outsourcing and increasing wage inequality

While many authors have studied the growth in wage inequality between employers and others have studied the impact of outsourcing on wages in particular occupations and industries, this paper is among the first to connect the two with a study of the impact of the changing distribution of occupations between employers on wage inequality in the United States. This paper uses multiple measures of occupational homogeneity (at both the establishment and employer tax-ID levels) to examine the impact of outsourcing on wages and on wage inequality. These measures show greater occupational homogeneity for the occupations used to study outsourcing by Abraham and Taylor (1996); Dube and Kaplan (2010); Dey, Houseman, and Polivka (2010); and Goldschmidt and Schmeider (2015), when these occupations are employed in establishments in the outsourced sector. For example, employer occupational homogeneity is higher for janitors when they are employed in establishments in the janitorial services industry than when they are employed in other industries.

The advantage of measuring outsourcing with occupational homogeneity is that these measures can be calculated for every employee of every employer, not only for "case study" occupations. This paper shows that by two very different measures of occupational homogeneity—for employers of every size—there is an increase in employer homogeneity over time for the quintile of workers in the lowest-wage occupations. Falling employment levels for middle-wage occupations mean those in other occupations have fewer coworkers in middle-wage occupations, but low-wage workers also have a declining share over time of coworkers in high-wage occupations, even as low-wage and high-wage occupations make up a growing share of employment. Low-wage occupations are growing in smaller employers, while the growth of high-wage occupations is concentrated in large employers. These patterns of time trends are consistent with the idea that in the economy as a whole, companies are "de-verticalizing" by outsourcing functions not part of their "core competencies," particularly if these outsourced tasks are done by workers paid lower wages than the "core workers" in the establishment.

The paper further shows that employer occupational homogeneity is related to wage levels. It has a particularly strong negative wage association for workers in occupations that are typically low paid, even after controlling for the occupations of employees and various observable characteristics of their employers. In contrast, workers in the highest paid quintile of occupations have a strong positive association with employer occupational homogeneity after controlling for their own occupations and the observable characteristics of their employers.

Song, Price, Guvenen, and von Wachter (2016) show that the vast majority of pay-inequality growth at small and medium-sized firms is due to the increasing segregation and sorting of workers who earn lower pay—without describing what about these workers makes them lower-paid workers—to firms that pay lower wages. Occupation is just such a characteristic affecting workers' wages, and this paper shows that workers in low-wage occupations are increasingly concentrated at employers with fewer high-wage occupations, contributing to wage inequality growth.

Although the data used in this paper cannot show changes in the wage distribution for the very highest 1.5% of wage-earners, they are well suited to measure the contribution of employers' occupational homogeneity to wage inequality growth for the remaining 98.5% of the wage distribution. Decompositions of ln(wage) variance growth in these data show very little role for the changing composition of employment by industry, geography, and establishment size. However, the growing polarization of employment can explain about half of inequality growth, and measures of the changing distribution of occupational homogeneity explain nearly all of the remainder. The form of occupational homogeneity that appears to matter most for growing inequality is the homogeneity of employment by the typical wage level of occupations. Growing separation of workers in low-wage occupations from the employers of workers in high-wage occupations is an important component of recent wage inequality growth.

REFERENCES

Abowd, John M., Francis Kramarz, and David Margolis (1999). “High Wage Workers and High Wage Firms.” *Econometrica*, pp. 251-334.

Abraham, Katharine G. and James R. Spletzer (2009). “New Evidence on the Returns to Job Skills.” *American Economic Review Papers and Proceedings*, Vol. 99, No. 2, pp. 52-57.

Abraham, Katharine G. and James R. Spletzer (2010). “Are the New Jobs Good Jobs?,” in *Labor in the New Economy*, edited by Katharine G. Abraham, James R. Spletzer, and Michael Harper. University of Chicago Press.

Abraham, Katherine G and Susan K. Taylor (1996). “Firms’ use of Outside Contractors: Theory and Evidence.” *Journal of Labor Economics*, 14(3): 394-424.

Acemoglu, Daron (2002). “Technical Change, Inequality, and the Labor Market.” *Journal of Economic Literature*, Vol. 40, No. 1, pp. 7-72.

Autor, David, H., and David Dorn (2013). “The growth of low-skill service jobs and the polarization of the US labor market.” *The American Economic Review*, Vol.103, No. 5, pp. 1553-1597.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney (2006). “The Polarization of the U.S. Labor Market,” *American Economic Association Annual Meeting Papers and Proceedings*, pp. 189-194.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney (2008). “Trends in U.S. Wage Inequality: Re-Assessing the Revisionists.” *Review of Economics and Statistics*, Vol. 90, No. 2, pp. 300-323.

Barth, Erling, Alex Bryson, James C. Davis, and Richard Freeman (2016). “It's Where You Work: Increases in Earnings Dispersion across Establishments and Individuals in the U.S.” *Journal of Labor Economics*, 34(S2): S67-S97.

Bound, John and George Johnson (1992). “Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations.” *The American Economic Review*, Vol. 82, No. 3, pp. 371-392.

Bronars, Stephen G. and Melissa Famulari (1997). “Wage, Tenure, and Wage Growth Variation Within and Across Establishments.” *Journal of Labor Economics*, pp. 285-317.

Card, David, Jorg Heining, and Patrick Kline (2013). “Workplace Heterogeneity and the Rise of West German Wage Inequality.” *Quarterly Journal of Economics*, Vol. 128, No. 3, pp. 967-1015.

Dey, Matthew, and Elizabeth Weber Handwerker (2016). "Longitudinal data from the Occupational Employment Statistics survey." *Monthly Labor Review*.

Dey, Matthew, Susan Houseman, and Anne Polivka. 2010. "What Do We Know about Contracting Out in the United States? Evidence from Household and Establishment Surveys." In Labor in the New Economy, Katharine G. Abraham, James R. Spletzer, and Michael J. Harper, eds. Chicago: University of Chicago Press, 2010, pp. 267-304

Davis, Steve J. and John Haltiwanger (1991). "Wage Dispersion Between and Within U.S. Manufacturing Plants." *Brookings Papers on Economic Activity*, pp. 115-200.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux (1996). "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica*, Vol. 64, No. 5, pp. 1001-1044.

Dube, Arindrajit, and Ethan Kaplan (2010). "Does outsourcing reduce wages in the low-wage service occupations? Evidence from janitors and guards." *ILR Review* 63.2: 287-306.

Dunne, Timothy, Lucia Foster, John Haltiwanger, and Kenneth R. Troske (2004). "Wage and Productivity Dispersion in United States Manufacturing: The Role of Computer Investment." *Journal of Labor Economics*, Vol. 22, No. 2, pp. 397-429.

Employee Benefit Research Institute (EBRI) (2009). *Fundamentals of Employee Benefit Programs, 6th ed.* http://www.ebri.org/publications/books/index.cfm?fa=fundamentals

Erickcek, George, Susan Houseman, and Arne Kalleberg (2003). "The Effects of Temporary Services and Contracting Out on Low-Skilled Workers: Evidence from Auto Suppliers, Hospitals, and Public Schools," in *Low-Wage America: How Employers are Reshaping Opportunity in the Workplace*, edited by Eileen Appelbaum, Annette Bernhardt, and Richard J. Murnane. New York: Russell Sage Foundation, pp.368-403.

Firpo, Sergio P., Nicole M. Fortin, and Thomas Lemieux (2018). "Decomposing Wage Distributions Using Recentered Influence Function Regressions." *Econometrics* 6 (3):41.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo (2011). "Decomposition Methods in Economics," *Handbook of labor economics* 4 (2011): 1-102.

Goldschmidt, Deborah and Johannes F. Schmieder (2015). "The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure," NBER Working Paper #21366. *The Quarterly Journal of Economics, forthcoming?*

Goos, Maarten and Alan Manning (2007). "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain," *Review of Economics and Statistics*, Vol. 89, No. 1, pp. 118-133.

Goos, Maarten, Alan Manning and Anna Salomons (2009). "Job Polarization in Europe," *American Economic Review Papers and Proceedings*, Vol. 99, No. 2, pp. 58-63.

Groshen, Erica L. (1991a). “Sources of Intra-Industry Wage Dispersion: How Much do Employers Matter?” *The Quarterly Journal of Economics*, Vol. 106, No. 3, pp. 869-884.

Groshen, Erica L. (1991b). “Five Reasons Why Wages Vary Among Employers.” *Industrial Relations*, Vol. 30, No. 3, pp. 350-381.

Handwerker, Elizabeth Weber, & Mason, Lowell G. (2013). Linking firms with establishments in BLS microdata. *Monthly Labor Review*, Vol 136, No 14.

Handwerker, Elizabeth Weber and James R. Spletzer (2014). “Measuring the distribution of wages in the United States from 1996-2010 with the Occupational Employment Survey” *Monthly Labor Review*, May 2014

Handwerker, Elizabeth Weber, and James R. Spletzer (2016). "The Role of Establishments and the Concentration of Occupations in Wage Inequality." *Inequality: Causes and Consequences*. Emerald Group Publishing Limited, 2016. 167-193.

Juhn, Chinhui, Kevin M. Murphy, Brooks Pierce (1993). “Wage Inequality and the Rise in Returns to Skill Wage Inequality and the Rise in Returns to Skill.” *The Journal of Political Economy*, Vol. 101, No. 3, pp. 410-442.

Katz, Lawrence F. and Autor, David H. (1999). “Changes in the Wage Structure and Earnings Inequality.” In *Handbook of Labor Economics* (volume 3A), edited by Orley Ashenfelter and David Card, Elsevier, pp. 1463-1555.

Katz, Lawrence F. and Kevin M. Murphy (1992). “Changes in Relative Wages, 1963-1987: Supply and Demand Factors.” *The Quarterly Journal of Economics*, Vol. 107, No. 1, pp. 35-78.

Lane, Julia. I., Salmon, Laurie A., & Spletzer, James R. (2007). Establishment wage differentials. *Monthly Labor Review*, Vol. 130,No.3.

Lemieux, Thomas (2006). “Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?” *American Economic Review*, Vol. 96, No. 3, pp. 461-498.

Lemieux, Thomas (2008). “The Changing Nature of Wage Inequality.” *Journal of Population Economics*, Vol. 21, No. 1, pp. 21-48.

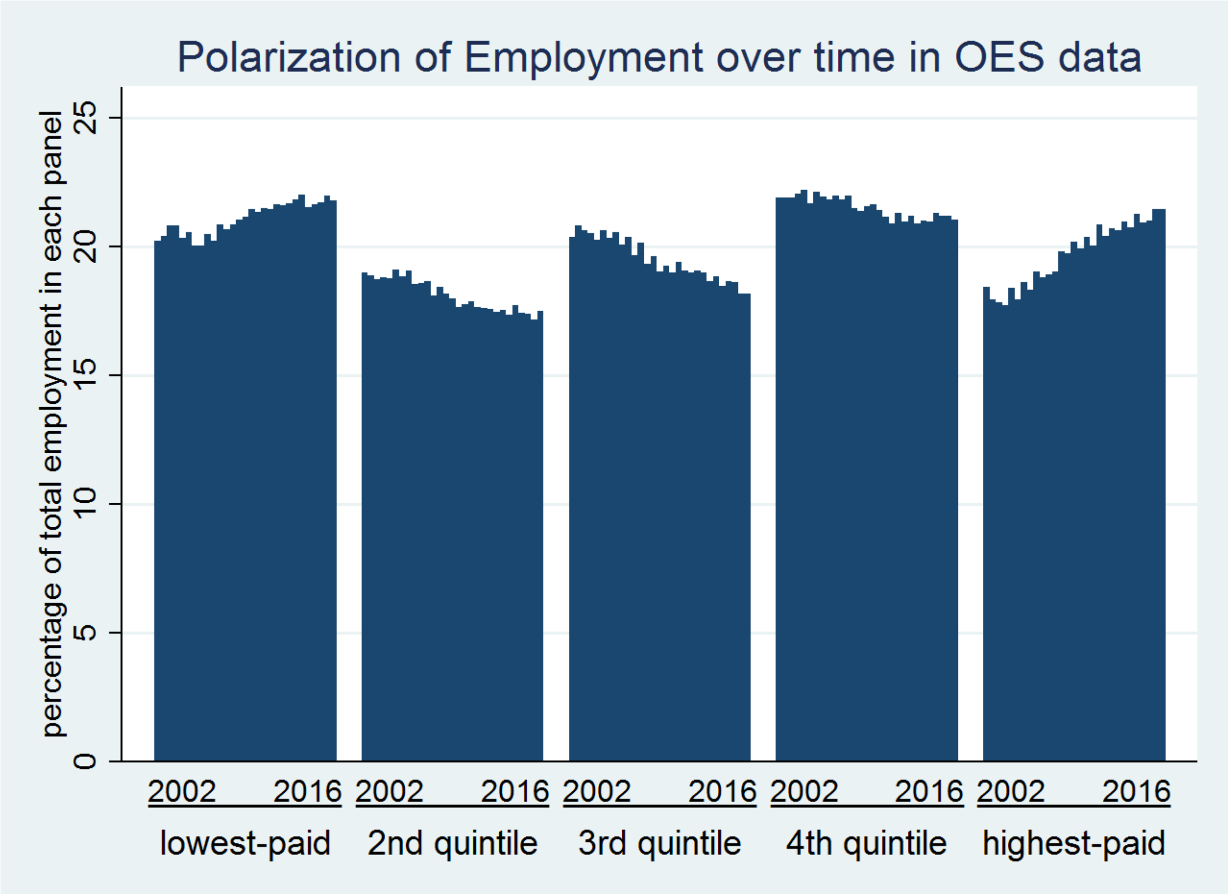
Perun, Pamela, "FAQs about Employees and Employee Benefits" (2010). *Georgetown Law School Memos and Fact Sheets*. Paper 46. Available at http://scholarship.law.georgetown.edu/legal/46

Rios-Avila, Fernando, “Recentered Influence Functions (RIF) in Stata: RIF Regression and RIF Decomposition” *Stata Journal*. Forthcoming.

Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. (2016) Firming Up Inequality. No. w21199. National Bureau of Economic Research, 2015.

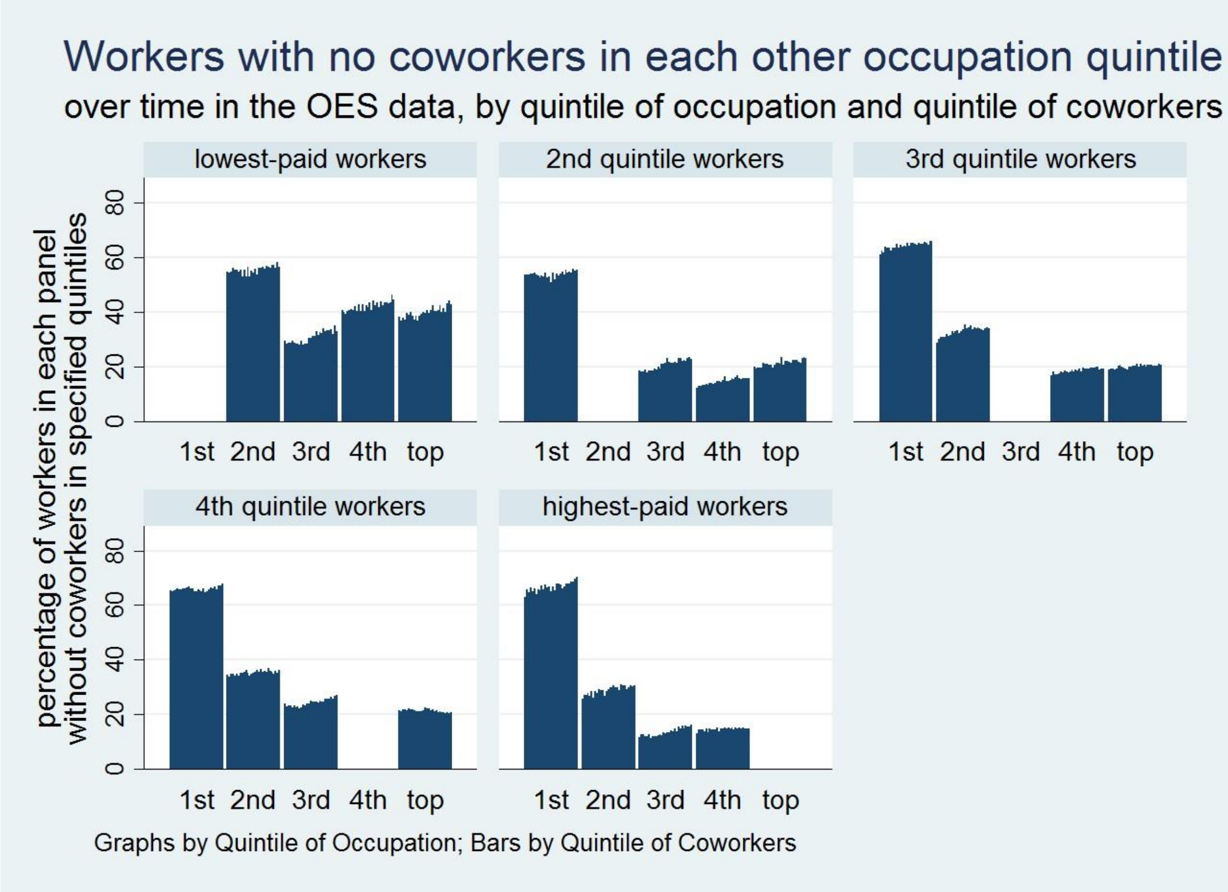
Weil, David. The fissured workplace. Harvard University Press, 2014.

Figure 1: Trends in Employment by Quintile of Occupation



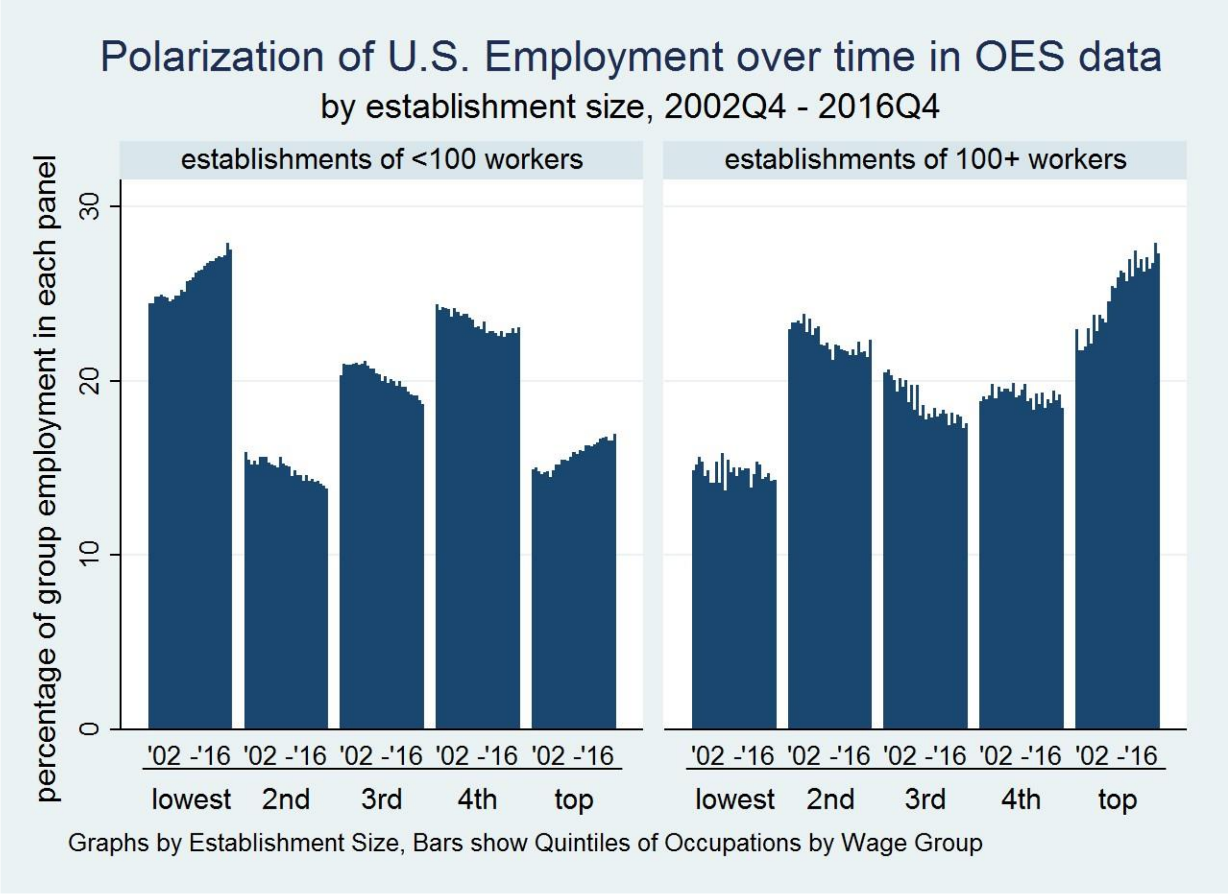
Note: The 46,609,394 observations in 29 waves of data are used to calculate overall average wage levels and employment levels. These are grouped into quintiles of occupation by average wage levels (as shown in Appendix A). This figure shows the percentage of employment in each occupational quintile in each wave of OES data, from November 2002 through November 2016.

Figure 2: Workers with no coworkers in other occupational quintiles over time in the OES data, by quintile of occupation and quintile of coworkers



Note: The 46,609,394 observations in 29 waves of data are used to calculate overall average wage levels and employment levels. These are grouped into quintiles of occupations by average wage levels (as shown in Appendix A). This figure shows the percentage of workers in each quintile who are employed in establishments that have no workers in each other quintile, by panel (from November 2002 to November 2016). For example, the subgraph at the top left shows the fraction of workers in the lowest-quintile of occupations who have no co-workers in each other quintile of occupations, for each panel of the OES data.

Figure 3: Trends in Employment by Quintile of Occupation and Size of Employing Establishment



Note: The 46,609,394 observations in 29 waves of data are used to calculate overall average wage levels and employment levels. These are grouped into quintiles of occupation by average wage levels (as shown in Appendix A). This figure shows the percentage of employment in each establishment size group in each occupational quintile in each wave of OES data, from November 2002 through November 2016.

Table 1: Mean Values of Occupational Homogeneity for Specified Occupations and Industries, 2002-2016

		Mean Value of Occupational Homogeneity	
Occupation and Industry	Avg ln(wage)	Herfindahl of Occupational Homogeneity for the establishment	Predicted Variance of Wages for the establishment
Food preparation and serving (SOC 35)			
within Food Services (NAICS 722) – 79%	1.99	0.496	0.138
within all other industries – 21%	2.11	0.252	0.245
Janitors (SOC 372011)			
within Janitorial Services (NAICS 561720) – 46%	2.04	0.842	0.141
within all other industries – 54%	2.16	0.329	0.267
Security Guards (SOC 339032)			
within Security Guard Srvcs (NAICS 561612) – 62%	2.16	0.883	0.158
within all other industries – 38%	2.32	0.322	0.267
Truck Drivers (SOC 53303)			
within Truck Transportation (NAICS 484) – 30%	2.69	0.636	0.204
within all other industries – 70%	2.45	0.379	0.248
Accountants (SOC 132011)			
within Accounting Services (NAICS 541211) – 25%	3.22	0.574	0.307
within all other industries – 75%	3.16	0.285	0.343
Computer Occupations (SOC 151)			
within Computer Services (NAICS 5415) – 27%	3.34	0.588	0.272
within all other industries – 73%	3.31	0.302	0.329
Engineers (SOC 172)			
within Engineering Services (NAICS 54133) – 22%	3.40	0.401	0.264
within all other industries – 78%	3.44	0.249	0.309
Lawyers (SOC 231011)			
within Law Offices (NAICS 54111) – 84%	3.76	0.411	0.544
within all other industries – 16%	3.87	0.277	0.409

Table 2: Change in mean values of Occupational Homogeneity over time

Trend regressions for quintiles of unimputed OES private-sector data from Nov 2002-Nov 2016

Occupations are ranked by average wage and classified into quintiles as shown in Appendix A

Occupational Homogeneity Variable	Herfindahl of occupational homogeneity for the establishment	Predicted Variance of Wages for the establishment
All Occupations (46,609,394 observations)		
Raw trend	0.0057	0.0059
	(0.00009)	(0.00004)
All Controls	0.0049	0.0048
	(0.00006)	(0.00003)
Lowest-paid quintile of occupations (4,512,045 observations)		
Raw trend	0.0153	-0.0167
	(0.00030)	(0.00011)
All Controls	0.0050	-0.0123
	(0.00020)	(0.00008)
Second quintile of occupations (6,742,713 observations)		
Raw trend	0.0134	-0.0009
	(0.00024)	(0.00009)
All Controls	0.0038	0.0032
	(0.00016)	(0.00007)
Middle quintile of occupations (9,757,141 observations)		
Raw trend	0.0073	0.0080
	(0.00017)	(0.00008)
All Controls	0.0011	0.0099
	(0.00013)	(0.00006)
Fourth quintile of occupations (11,881,704 observations)		
Raw trend	-0.0079	0.0140
	(0.00018)	(0.00006)
All Controls	0.0027	0.0102
	(0.00013)	(0.00004)
Highest-paid quintile of occupations (13,715,791 observations)		
Raw trend	0.0065	0.0187
	(0.00014)	(0.00006)
All Controls	0.0101	0.0157
	(0.00011)	(0.00004)

Note: These are coefficients α from regressions of the form $Occupation\ Homogeneity_{ijt} = \alpha Survey\ Date_t + \beta I(May\ Survey_t) + \gamma X_{ijt} + \varepsilon_{ijt}$, where the Survey Date is measured in decades since X_{ijt} includes occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, establishment size class, continuous establishment size, and state. All coefficients are significant at $p < 0.001$. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Standard errors in parentheses

Table 3: Change in mean values of Occupational Homogeneity over time, by establishment size

Trend regressions for quintiles of unimputed OES private-sector data from Nov 2002-Nov 2016
Occupations are ranked by average wage and classified into quintiles as shown in Appendix A

	Occupational Homogeneity Variable		Herfindahl of occupational homogeneity for the establishment	Predicted Variance of Wages for the establishment
1-99 Employees				
	All Occupations (27,077,995 observations)			
		All Controls	0.0091	0.0034
			(0.0001)	(0.0000)
	Lowest-paid quintile of occupations (2,942,146 observations)			
		All Controls	0.0120	-0.0137
			(0.0003)	(0.0001)
100+ Employees				
	All Occupations (19,531,399 observations)			
		All Controls	0.0035	0.0066
			(0.0001)	(0.0000)
	Lowest-paid quintile of occupations (1,569,899 observations)			
		All Controls	-0.0057	-0.0100
			(0.0003)	(0.0001)

Note: These are coefficients α from regressions of the form $Occupation\ Homogeneity_{ijt} = \alpha Survey\ Date_t + \beta I(May\ Survey_t) + \gamma X_{ijt} + \varepsilon_{ijt}$, where the Survey Date is measured in decades since X_{ijt} includes occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, establishment size class, continuous establishment size, and state. All coefficients are significant at $p < 0.001$. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Standard errors in parentheses

Table 4: Regressions of log wages on measures of Occupational Homogeneity

Wage regressions for quintiles of unimputed OES private-sector data from Nov 2002 - Nov 2016

Occupations are ranked by average wage and classified into quintiles as shown in Appendix A

Occupational Homogeneity Variable	Herfindahl of occupational homogeneity for the establishment	Predicted Variance of Wages for the establishment
All Occupations (46,609,394 observations)		
With only date fixed effects	-0.496	2.285
	(0.001)	(0.002)
All Controls	-0.107	0.234
	(0.000)	(0.001)
Lowest-paid quintile of occupations (4,512,045 observations)		
With only date fixed effects	-0.229	0.882
	(0.001)	(0.003)
All Controls	-0.145	0.358
	(0.001)	(0.003)
Second quintile of occupations (6,742,713 observations)		
With only date fixed effects	-0.254	0.722
	(0.001)	(0.003)
All Controls	-0.159	0.377
	(0.001)	(0.002)
Middle quintile of occupations (9,757,141 observations)		
With only date fixed effects	-0.199	0.579
	(0.001)	(0.002)
All Controls	-0.122	0.295
	(0.001)	(0.002)
Fourth quintile of occupations (11,881,704 observations)		
With only date fixed effects	-0.249	0.647
	(0.001)	(0.003)
All Controls	-0.074	0.096
	(0.001)	(0.003)
Highest-paid quintile of occupations (13,715,791 observations)		
With only date fixed effects	-0.206	0.463
	(0.001)	(0.003)
All Controls	0.009	-0.153
	(0.001)	(0.003)

Notes: These are coefficients α from regressions of the form $\ln(wage_{ijt}) = \alpha OccHomogeneity_{jt} + \beta OccConcen_{jt} Date_t + \gamma X_{ijt} + \varepsilon_{ijt}$, where Date is measured in decades since November 1, 2002, X includes survey date fixed effects, occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, state fixed effects, and establishment size (using fixed effects for establishment size classes as well as continuous establishment size). All coefficients are significant at $p < 0.001$. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Standard errors in parentheses.

Table 5: Decomposition of Changes in real ln(wage) variance from 2002-2003 to 2014-2015

real log wage variance	Coeff.	Percent	Std. Err.	z	P>|z|	[95% Conf. Interval]	
Overall Variance							
Late period (2014-2015)	0.3999		0.0013	315.24	0	0.3974	0.4024
counterfactual variance	0.3724		0.0010	375.45	0	0.3705	0.3744
Early period (2002-2003)	0.3754		0.0016	233.34	0	0.3722	0.3786
Total change	0.0245	100%	0.0020	11.95	0	0.0205	0.0285
of which explained (by compositional change)	0.0274	112%	0.0005	55.88	0	0.0265	0.0284
of which unexplained (wage structure change)	-0.0030	-12%	0.0019	-1.55	0.122	-0.0067	0.0008
Explained (compositional effect)							
Total	0.0274	100%	0.0005	55.88	0	0.0265	0.0284
Pure_explained	0.0268	98%	0.0005	49.91	0	0.0258	0.0279
Specification_error	0.0006	2%	0.0001	5.53	0	0.0004	0.0008
Components of the pure explained effect							
industry	-0.0002	-1%	0.0002	-1.11	0.267	-0.0006	0.0002
geography	-0.0001	0%	0.0001	-1.32	0.186	-0.0003	0.0001
size	0.0005	2%	0.0001	6.05	0	0.0004	0.0007
quintiles of occupation	0.0140	51%	0.0005	26.66	0	0.0129	0.0150
Establishment Herfindahls	0.0015	6%	0.0003	5.51	0	0.0010	0.0021
Establishment predicted var(ln wages)	0.0111	41%	0.0006	19.21	0	0.0100	0.0123
Unexplained (wage structure changes)							
Total	-0.0030	100%	0.0019	-1.55	0.122	-0.0067	0.0008
Reweight_error	-0.0001	4%	0.0008	-0.13	0.898	-0.0017	0.0015
Pure_Unexplained	-0.0029	96%	0.0017	-1.65	0.099	-0.0063	0.0005

Notes: These are the results of Rios-Avila's implementation of Fortin, Firpo, and Lemiueux's Recentered Influence Function Decomposition. Industry here is grouped into 19 supersectors and geography into 7 Census divisions. Establishment size is measured in 8 categories (1-4 employees, 5-9, 10-19, 20-49, 50-99, 100-249, 250-499, and 500+). Establishment-level Herfindahl-Hirschman indices of occupations are measured with quartiles of the distribution, interacted with occupational quintiles. The predicted variance of ln(wage) for establishments is divided into quartiles, and also interacted with occupational quintiles, with an additional dummy variable for low-wage occupations in establishment of less than 100 workers that are in the bottom half of the predicted variance distribution. Standard errors presented here have not been bootstrapped—the bootstrap standard errors are still processing and should be available by late summer 2020 at current computer speeds. Additional specifications (without bootstrapped standard errors) are shown in Appendix C and Appendix D.

Data Appendix

This paper uses Occupational Employment Statistics (OES) Survey microdata. The OES survey is designed to measure occupational employment and wages in the United States by geography and industry, and is the only such survey of its size and scope, covering all establishments in the United States except those in agriculture, private households, and unincorporated self-employed workers without employees. Every year, approximately 400,000 private and local government establishments are asked to report the number of employees in each occupation paid within specific wage intervals: 200,000 establishments each November and another 200,000 each May. As described in Dey and Handwerker, the OES uses a complex sample design intended to minimize the variance of wage estimates for each occupation within industries and geographic areas. Thus, establishments expected to employ occupations with greater variation in wages have relatively larger probabilities of selection and lower estimation weights.

The OES survey form is a matrix of detailed occupations and wage intervals. For large establishments, the survey form lists 50 to 225 detailed occupations; these occupations pre-printed on the survey form are selected based on the industry and the size of the establishment. Small establishments write descriptions of the work done by their employees, which are coded into occupations by staff in state labor agencies. Wage intervals on the OES survey form are given in both hourly and annual nominal dollars, with annual earnings that are 2080 times the hourly wage rates. To calculate average wages, the OES program obtains the mean of each wage interval every year from the National Compensation Survey (NCS). These mean wages are then assigned to all employees in that wage interval. The OES survey is not designed to produce time series statistics. Time series in this paper are produced using the methodology described in Abraham and Spletzer (2010) to reweight the data to November or May benchmarks of total employment by detailed industry and by broad industry and establishment size groups from the Quarterly Census of Employment and Wages (QCEW).

The OES has been using the Standard Occupational Classification System since 1999, and had a change of industry classification systems from SIC to NAICS (2002) soon thereafter. Certain SOC and NAICS codes are combined to make groups consistent across the 2007 and 2012 NAICS revisions and the 2010 revision to the SOC. Data used in this paper begin in 2002 to avoid inconsistencies of SOC coding in small establishments during the initial years that the OES program used this coding system, as described by Abraham and Spletzer (2010).

Handwerker and Spletzer (2016) examine the decomposition of total wage variance in the OES into its within-establishment and between establishment components at length. Updating their findings, over the period of Fall 1999 through November 2016, 60% of wage variance is between establishments, while all of the growth in overall wage variance over this period is between establishments. Handwerker and Spletzer (2016) also find that similar amounts of establishment-level wage variance in the OES can be explained by broad industry groups to the amount found by Barth, Bryson, Davis, and Freeman. However, more of the establishment-level wage variance can be explained by detailed industry in the OES data than in the Census data, echoing findings comparing OES and CPS data.

Data Appendix Table 1: Summary Statistics

Variable description	Occupation by wage interval observations	Employment represented	Weighted Mean	Minimum	Maximum	Variance
OES wages, by establishment-occupation-wage interval	46,609,428	2,242,528,409	20.043	5.778	145.759	310.785
OES real wages, by estab-occupation-wage interval	46,609,428	2,242,528,409	16.203	5.209	106.443	197.447
OES real ln wage, by estab-occupation-wage interval	46,609,428	2,242,528,409	2.561	1.650	4.668	0.382
measured var(ln(wg)) of establishment	46,609,428	2,242,528,409	0.154	0.000	2.091	0.019
Herfindahl of 3-digit occupations, establishment level	46,609,428	2,242,528,409	0.408	0.032	1.000	0.063
pred var(ln(wg)) for establishment, based on occupations	46,609,394	2,242,528,409	0.270	0.018	1.016	0.012
portion of above due to variation between occupations	46,609,394	2,242,528,409	0.106	0.000	0.781	0.006
predicted mean(ln(wg)) of estab, based on occupations	46,609,394	2,242,528,409	2.572	1.807	3.851	0.113
Total employment of establishment	46,609,428	2,242,528,409	576	1	56,473	5,104,405
Date of observation	46,609,428	2,242,528,409	Nov, 2009	Nov, 2002	Nov, 2016	

Variable description	Occupation by wage interval observations	Employment represented	Fraction of employment	Establishment observations
Quintiles				
Bottom quintile of occupations	4,512,048	498,610,022	22.2%	1,298,634
Second quintile of occupations	6,742,720	404,179,707	18.0%	1,946,156
Third quintile of occupations	9,757,150	434,006,537	19.4%	2,515,660
Fourth quintile of occupations	11,881,718	486,554,202	21.7%	2,806,055
Top quintile of occupations	13,715,792	419,177,942	18.7%	2,335,591
Industries (2-digit)				
Agriculture, Forestry, Fishing and Hunting	24,742	1,209,745	0.1%	5,064
Mining, Quarrying, and Oil and Gas Extraction	278,611	10,731,557	0.5%	22,949
Utilities	362,131	11,487,328	0.5%	20,980
Construction	2,715,765	140,698,009	6.3%	320,008
Manufacturing	7,617,106	247,417,992	11.0%	404,392
Wholesale Trade	3,121,983	106,037,580	4.7%	289,829
Retail Trade	6,033,799	344,932,292	15.4%	502,403
Transportation and Warehousing	1,372,852	92,123,403	4.1%	142,832
Information	1,509,864	49,886,150	2.2%	110,529
Finance and Insurance	2,691,456	118,883,973	5.3%	227,072
Real Estate and Rental and Leasing	673,419	28,956,088	1.3%	99,687
Professional, Scientific, and Technical Services	3,514,547	143,457,542	6.4%	354,483
Management of Companies and Enterprises	1,276,951	32,276,040	1.4%	38,583
Administrative and Support and Waste Management and	2,291,647	140,897,584	6.3%	246,780
Educational Services	1,298,489	50,396,572	2.2%	70,057
Health Care and Social Assistance	6,809,349	353,872,634	15.8%	420,820
Arts, Entertainment, and Recreation	1,062,174	37,963,004	1.7%	96,342
Accommodation and Food Services	2,078,374	237,573,956	10.6%	209,645
Other Services (except Public Administration)	1,876,169	93,726,962	4.2%	282,115
Occupations (2-digit)				
Management Occupations	5,121,383	103,271,611	4.6%	
Business and Financial Operations Occupations	3,695,454	99,137,164	4.4%	
Computer and Mathematical Occupations	1,705,143	53,350,145	2.4%	
Architecture and Engineering Occupations	1,210,073	39,732,902	1.8%	
Life, Physical, and Social Science Occupations	347,605	12,029,141	0.5%	
Community and Social Service Occupations	611,383	23,933,901	1.1%	
Legal Occupations	249,027	16,324,258	0.7%	
Education, Training, and Library Occupations	654,815	38,966,650	1.7%	
Arts, Design, Entertainment, Sports, and Media Occupatio	981,616	28,832,089	1.3%	
Healthcare Practitioners and Technical Occupations	2,195,949	134,205,129	6.0%	
Healthcare Support Occupations	732,683	75,368,512	3.4%	
Protective Service Occupations	322,367	23,877,442	1.1%	
Food Preparation and Serving Related Occupations	1,918,555	223,615,488	10.0%	
Building and Grounds Cleaning and Maintenance	1,125,051	72,733,714	3.2%	
Personal Care and Service Occupations	800,258	73,439,546	3.3%	
Sales and Related Occupations	4,545,163	297,008,580	13.2%	
Office and Administrative Support Occupations	10,417,328	375,746,893	16.8%	
Farming, Fishing, and Forestry Occupations	63,893	3,315,741	0.1%	
Construction and Extraction Occupations	1,546,401	108,571,047	4.8%	
Installation, Maintenance, and Repair Occupations	2,384,025	97,043,195	4.3%	
Production Occupations	3,467,582	174,526,262	7.8%	
Transportation and Material Moving Occupations	2,513,674	167,498,997	7.5%	

Appendix A: Occupations by Quintile

3-digit SOC code	*SOC Title*	*Average ln(wage)*	*Cummulative percentage of employment*	*Occupation Quintile*
359	Other Food Preparation and Serving Related Workers	1.94	1.1%	1
353	Food and Beverage Serving Workers	1.94	6.5%	1
393	Entertainment Attendants and Related Workers	1.98	6.9%	1
352	Cooks and Food Preparation Workers	2.05	9.3%	1
392	Animal Care and Service Workers	2.09	9.4%	1
399	Other Personal Care and Service Workers	2.09	11.1%	1
412	Retail Sales Workers	2.10	18.6%	1
372	Building Cleaning and Pest Control Workers	2.10	20.9%	1
452	Agricultural Workers	2.11	21.0%	2
536	Other Transportation Workers	2.13	21.2%	2
516	Textile Apparel and Furnishings Workers	2.15	21.9%	2
311	Nursing Psychiatric and Home Health Aides	2.16	23.8%	2
396	Baggage Porters Bellhops and Concierges	2.19	23.9%	2
395	Personal Appearance Workers	2.20	24.3%	2
373	Grounds Maintenance Workers	2.21	25.0%	2
339	Other Protective Service Workers	2.23	26.0%	2
397	Tour and Travel Guides	2.23	26.0%	2
513	Food Processing Workers	2.24	26.7%	2
537	Material Moving Workers	2.24	30.6%	2
379	Other Buildings, Grounds, and Maintenance Occupations	2.27	30.6%	2
259	Other Education Training and Library Occupations	2.28	30.9%	2
435	Material Recording Scheduling Dispatching and	2.29	33.8%	2
432	Communications Equipment Operators	2.29	34.0%	2
473	Helpers Construction Trades	2.31	34.2%	2
459	Other Farming, Fishing, and Forestry Occupations	2.32	34.2%	2
453	Fishing and Hunting Workers	2.33	34.2%	2
439	Other Office and Administrative Support Workers	2.35	37.3%	2
517	Woodworkers	2.36	37.5%	2
512	Assemblers and Fabricators	2.40	39.2%	2
434	Information and Record Clerks	2.40	43.5%	3
519	Other Production Occupations	2.40	45.9%	3
319	Other Healthcare Support Occupations	2.42	47.0%	3
351	Supervisors of Food Preparation and Serving Workers	2.45	47.7%	3
433	Financial Clerks	2.48	50.5%	3
332	Fire Fighting and Prevention Workers	2.48	50.5%	3
533	Motor Vehicle Operators	2.49	53.5%	3
454	Forest Conservation and Logging Workers	2.50	53.5%	3
515	Printing Workers	2.53	53.8%	3
219	Other Community and Social Service Occupations	2.53	53.8%	3

3-digit SOC code	*SOC Title*	*Average ln(wage)*	*Cummulative percentage of employment*	*Occupation Quintile*
252	Preschool Primary Secondary and Special Education	2.54	54.4%	3
514	Metal Workers and Plastic Workers	2.56	56.2%	3
394	Funeral Service Workers	2.56	56.3%	3
436	Secretaries and Administrative Assistants	2.57	59.0%	3
419	Other Sales and Related Workers	2.57	59.8%	4
253	Other Teachers and Instructors	2.58	60.1%	4
333	Law Enforcement Workers	2.58	60.1%	4
391	Supervisors of Personal Care and Service Workers	2.59	60.2%	4
211	Counselors Social Workers and Other Community and	2.59	61.2%	4
371	Supervisors of Building and Grounds Cleaning and	2.62	61.4%	4
493	Vehicle and Mobile Equipment Mechanics Installers	2.63	62.6%	4
312	Occupational Therapy and Physical Therapist Assistants	2.65	62.8%	4
499	Other Installation Maintenance and Repair Occupations	2.65	64.9%	4
212	Religious Workers	2.66	64.9%	4
534	Rail Transportation Workers	2.67	65.0%	4
475	Extraction Workers	2.68	65.1%	4
474	Other Construction and Related Workers	2.70	65.3%	4
292	Health Technologists and Technicians	2.71	67.4%	4
472	Construction Trades Workers	2.71	71.1%	4
274	Media and Communication Equipment Workers	2.72	71.2%	4
271	Art and Design Workers	2.77	71.7%	4
331	Supervisors of Protective Service Workers	2.77	71.7%	4
194	Life Physical and Social Science Technicians	2.77	71.9%	4
411	Supervisors of Sales Workers	2.78	73.2%	4
254	Librarians Curators and Archivists	2.79	73.2%	4
272	Entertainers and Performers Sports and Related	2.80	73.6%	4
451	Supervisors of Farming Fishing and Forestry Workers	2.81	73.6%	4
492	Electrical and Electronic Equipment Mechanics Install	2.81	74.1%	4
232	Legal Support Workers	2.83	74.4%	4
531	Supervisors of Transportation and Material Moving	2.87	74.7%	4
239	Other Legal Occupations	2.88	74.7%	4
431	Supervisors of Office and Administrative Support	2.90	75.8%	4
535	Water Transportation Workers	2.90	75.8%	4
299	Other Healthcare Practitioners and Technical Occs	2.91	75.9%	4
173	Drafters Engineering Technicians and Mapping	2.91	76.5%	4
273	Media and Communication Workers	2.93	77.0%	4
511	Supervisors of Production Workers	2.98	77.5%	4
413	Sales Representatives Services	3.01	78.9%	4
518	Plant and System Operators	3.01	79.0%	4
153	All other Computer and Math Occupations	3.05	79.0%	4

3-digit SOC code	SOC Title	Average ln(wage)	Cummulative percentage of employment	Occupation Quintile
414	Sales Representatives Wholesale and Manufacturing	3.08	80.7%	4
491	Supervisors of Installation Maintenance and Repair	3.08	81.0%	5
131	Business Operations Specialists	3.11	83.8%	5
471	Supervisors of Construction and Extraction Workers	3.11	84.2%	5
193	Social Scientists and Related Workers	3.16	84.3%	5
171	Architects Surveyors and Cartographers	3.17	84.4%	5
132	Financial Specialists	3.17	86.3%	5
251	Postsecondary Teachers	3.18	86.8%	5
159	Computer and Math Occupations, NEC	3.24	86.8%	5
532	Air Transportation Workers	3.26	87.0%	5
192	Physical Scientists	3.31	87.2%	5
151	Computer Specialists	3.32	89.8%	5
191	Life Scientists	3.32	90.0%	5
119	Other Management Occupations	3.33	91.3%	5
152	Mathematical Science Occupations	3.35	91.4%	5
291	Health Diagnosing and Treating Practitioners	3.39	94.8%	5
172	Engineers	3.43	96.0%	5
113	Operations Specialties Managers	3.60	97.2%	5
111	Top Executives	3.64	99.0%	5
112	Advertising Marketing Promotions Public Relations &	3.65	99.6%	5
231	Lawyers Judges and Related Workers	3.76	100.0%	5

Appendix B: Heterogeneity of Wage and Trend Results

B1: Heterogeneity by state-level unionization rates

One factor which may impact both wages and the organization of production in terms of the variety of occupations at a workplace is unionization. The OES does not collect information on unionization patterns by employer, but it includes location of each establishment, and unionization rates vary strongly by state. Thus, state-level unionization rates are used to group the data into highly unionized states (17-26% of employed workers unionized), middle, and low unionized states (3-9.3% unionized), based on published tables from the Current Population Survey.

Overall, workers in states with higher unionization levels work in slightly (but statistically significantly) less occupationally homogeneous establishments. However, for workers in the lowest-paid quintile of occupations, this reverses; these workers have slightly higher occupational homogeneity in establishments located in states with higher unionization levels.

Differences in occupational homogeneity trends between less and more unionized states show that establishments are growing more occupationally homogeneous over time in the less-unionized states, relative to the highly unionized states, by every measure. Following equation (3), occupational homogeneity as measured by the Herfindahl-Hirschman index appears to be increasing slightly (but significantly) faster in less-unionized states than in highly unionized states, both across all occupations and for occupations in the bottom quintile, and so workers in less unionized states now work in establishments with higher Herfindahl-Hirschman indices, on average, than workers in more unionized states. Occupational homogeneity as measured by the predicted variance of ln(wages) based on the occupational composition of establishments, and its between-occupations component also show less of an increase in the less unionized states, and after employer characteristics and occupation controls are included in equation (3), all measures of occupational homogeneity show statistically significant trends of increasing employer homogeneity in less unionized states. Differences in trends are similar for the lowest-paid quintile of occupations.

Following equation (4), the relationships between occupational homogeneity and wages are estimated separately for each unionization group of states. Across all occupations, the relationships α between occupational homogeneity (by all measures) and wages is significantly greater in the more highly unionized states. However, this reverses when establishment characteristics and occupational controls are included in equation (2), and after including these controls, the relationship between occupational homogeneity and wages (α) is significantly greater in the less unionized states than in the more highly unionized states. For workers in the lowest-paid quintile of occupations, occupational homogeneity (by all measures) matters more for wages in less unionized states both with and without controlling for establishment characteristics and occupation. Across all workers, time trends interacted with occupational homogeneity (β) have varying signs across measures of occupational homogeneity and the inclusion of controls, but for the workers in the lowest-paid quintile of occupations, these interactions are always significantly lower in less unionized states, indicating that the relationships between occupational homogeneity and wages for these workers are converging over time between the different groups of states.

Table B1: Change in mean values of Occupational Homogeneity over time, by Unionization group

Trend regressions for quintiles of unimputed OES private-sector data from Nov 2002-Nov 2016
Occupations are ranked by average wage and classified into quintiles as shown in Appendix A

Occupational Homogeneity Variable	Herfindahl of occupational homogeneity for the establishment	Predicted Variance of Wages for the establishment
Highly unionized states		
All Occupations (15,699,964 observations)		
Raw trend	0.0039	0.0096
	(0.0002)	(0.0001)
All Controls	0.0060	0.0086
	(0.0001)	(0.0000)
Lowest-paid quintile of occupations (1,494,981 observations)		
Raw trend	0.0095	(0.0135)
	(0.0005)	(0.0002)
All Controls	0.0042	(0.0085)
	(0.0003)	(0.0001)
Less unionized states		
All Occupations (16,120,909 observations)		
Raw trend	0.0097	0.0004
	(0.0001)	(0.0001)
All Controls	0.0090	(0.0012)
	(0.0001)	(0.0000)
Lowest-paid quintile of occupations (1,579,134 observations)		
Raw trend	0.0186	(0.0230)
	(0.0005)	(0.0002)
All Controls	0.0131	(0.0197)
	(0.0003)	(0.0001)

Note: These are coefficients α from regressions of the form $Occupational\ Homogeneity_{ijt} = \alpha Survey\ Date_t + \beta I(May\ Survey_t) + \gamma X_{ijt} + \varepsilon_{ijt}$, where the Survey Date is measured in decades since X_{ijt} includes occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, establishment size class, continuous establishment size, and state. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Standard errors in parentheses. Regressions at the establishment-occupation-wage interval level, weighted by employment.

Table B2: Regressions of log wages on measures of Occupational Homogeneity, by Unionization group

Wage regressions for quintiles of unimputed OES private-sector data from Nov 2002 - Nov 2016
Occupations are ranked by average wage and classified into quintiles as shown in Appendix A

Occupational Homogeneity Variable	Herfindahl of occupational homogeneity for the establishment	Predicted Variance of Wages for the establishment
Highly unionized states		
All Occupations (15,699,964 observations)		
With only date fixed effects	-0.494	2.326
	(0.001)	(0.003)
All Controls	-0.102	0.241
	(0.001)	(0.002)
Lowest-paid quintile of occupations (1,494,981 observations)		
With only date fixed effects	-0.186	0.817
	(0.002)	(0.005)
All Controls	-0.125	0.338
	(0.002)	(0.005)
Less unionized states		
All Occupations (16,120,909 observations)		
With only date fixed effects	-0.498	2.198
	(0.001)	(0.003)
All Controls	-0.134	0.260
	(0.001)	(0.002)
Lowest-paid quintile of occupations (1,579,134 observations)		
With only date fixed effects	-0.248	0.895
	(0.002)	(0.005)
All Controls	-0.166	0.377
	(0.002)	(0.004)

Notes: These are coefficients α from regressions of the form $\ln(wage_{ijt}) = \alpha Occupational\ Homogeneity_{jt} + \beta OccConcen_{jt} Date_t + \gamma X_{ijt} + \varepsilon_{ijt}$, where Date is measured in decades since November 1, 2002, X includes survey date fixed effects, occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, state fixed effects, and establishment size (using fixed effects for establishment size classes as well as continuous establishment size). All coefficients are significant at $p < 0.001$. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Standard errors in parentheses.

B2: Heterogeneity by establishment age

Because the Occupational Establishment Survey data is sampled from the records of the BLS Quarterly Census of Employment and Wages, which BLS assembles into a longitudinal database of establishments, it is straightforward to link these datasets together and find a "birth date"—the first quarter with employment greater than zero—for each establishment. Dividing establishments into those born before the fourth quarter of 2002 (87% of all establishments observed) and those born afterwards (13%), the analyses above can be repeated separately for "old" and "young" establishments.

Young establishments are, in employment-weighted averages, more homogeneous in occupations than old establishments, with higher Herfindahl indices of occupational homogeneity (.511 for young establishments compared with .386 for old establishments), and lower predicted variances of wages based on the occupational composition of establishments (.241 for young establishments compared with .276 for old establishments), with much of the difference due to the between-occupations component of this variance (.081 for young establishments compared with .111 for old establishments). This pattern is echoed at higher levels of employer homogeneity for the establishments of workers in the lowest-paid quintile of occupations—for these workers as well, working in younger establishments means working in establishments more occupationally homogeneous, by both measures.

There are no clear patterns in differences in trends in occupational homogeneity by establishment age—differences vary greatly by which measure of occupational homogeneity is used, whether controls for establishment characteristics are included, and which groups of occupations are examined.

Examining the relationships between occupational homogeneity and wages by establishment age, occupational homogeneity matters more for wages in old establishments than in young establishments than in young establishments. The difference between old and young establishments becomes much smaller after additional controls are added to equation 2, but occupational homogeneity—by every measure—still matters significantly more in old establishments than in young establishments. This is true across all workers as well as for workers in the lowest-paid quintile of occupations.

Table B3: Change in mean values of Occupational Homogeneity over time, by Establishment age group

Trend regressions for quintiles of unimputed OES private-sector data from Nov 2002-Nov 2016
Occupations are ranked by average wage and classified into quintiles as shown in Appendix A

Occupational Homogeneity Variable	Herfindahl of occupational homogeneity for the establishment	Predicted Variance of Wages for the establishment
Old Establishments		
All Occupations (6,119,160 observations)		
Raw trend	-0.0087	0.0117
	(0.00032)	(0.00013)
All Controls	-0.0025	0.0086
	(0.00024)	(0.00009)
Lowest-paid quintile of occupations (742,133 observations)		
Raw trend	0.0219	-0.0034
	(0.00090)	(0.00031)
All Controls	-0.0003	-0.0020
	(0.00062)	(0.00026)
Young establishments		
All Occupations (39,588,761 observations)		
Raw trend	-0.0261	0.0150
	(0.00009)	(0.00004)
All Controls	-0.0012	0.0060
	(0.00007)	(0.00003)
Lowest-paid quintile of occupations (3,659,326 observations)		
Raw trend	-0.0133	-0.0082
	(0.00033)	(0.00012)
All Controls	0.0019	-0.0122
	(0.00022)	(0.00009)

Note: These are coefficients α from regressions of the form $Occupational\ Homogeneity_{ijt} = \alpha Survey\ Date_t + \beta I(May\ Survey_t) + \gamma X_{ijt} + \varepsilon_{ijt}$, where the Survey Date is measured in decades since X_{ijt} includes occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, establishment size class, continuous establishment size, and state. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Standard errors in parentheses Regressions at the establishment-occupation-wage interval level, weighted by employment.

Table B4: Regressions of log wages on measures of Occupational Homogeneity, by Establishment age group

Wage regressions for quintiles of unimputed OES private-sector data from Nov 2002 - Nov 2016
Occupations are ranked by average wage and classified into quintiles as shown in Appendix A

Occupational Homogeneity Variable	Herfindahl of occupational homogeneity for the establishment	Predicted Variance of Wages for the establishment
Old Establishments		
All Occupations (6,119,160 observations)		
With only date fixed effects	-0.285	2.330
	(0.001)	(0.002)
All Controls	-0.104	0.167
	(0.001)	(0.002)
Lowest-paid quintile of occupations (742,133 observations)		
With only date fixed effects	-0.110	0.469
	(0.001)	(0.003)
All Controls	-0.097	0.187
	(0.002)	(0.004)
Young establishments		
All Occupations (39,588,761 observations)		
With only date fixed effects	-0.572	2.382
	(0.000)	(0.001)
All Controls	-0.099	0.201
	(0.000)	(0.001)
Lowest-paid quintile of occupations (3,659,326 observations)		
With only date fixed effects	-0.234	0.783
	(0.001)	(0.002)
All Controls	-0.131	0.227
	(0.001)	(0.002)

Notes: These are coefficients α from regressions of the form $\ln(wage_{ijt}) = \alpha Occupational\ Homogeneity_{jt} + \beta OccConcen_{jt} Date_t + \gamma X_{ijt} + \varepsilon_{ijt}$, where Date is measured in decades since November 1, 2002, X includes survey date fixed effects, occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, state fixed effects, and establishment size (using fixed effects for establishment size classes as well as continuous establishment size). All coefficients are significant at $p < 0.001$. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Standard errors in parentheses.

B3: Heterogeneity by establishment size

Subdividing establishments into the same size classes used in the regression controls (1-4 employees, 5-9, 10-19, 20-49, 50-99, 100-249, 250-499, and 500+), there is a mostly monotonic decrease in employment-weighted means of occupational homogeneity with establishment size, by every measure of occupational homogeneity. Larger establishments are more heterogeneous in occupations. There is no clear pattern across establishment sizes of relationships between occupational homogeneity and wages for all workers. However, for the lowest-paid quintile of occupations, the strongest relationships between occupational homogeneity and wages—across all measures, with and without controlling for other observable characteristics—appears in middle-sized establishments: those with 50-99 employees.

Table B5: Change in mean values of Occupational Homogeneity over time, by Establishment size group

Trend regressions for quintiles of unimputed OES private-sector data from Nov 2002-Nov 2016
Occupations are ranked by average wage and classified into quintiles as shown in Appendix A

	Occupational Homogeneity Variable	Herfindahl of occupational homogeneity for the establishment	Predicted Variance of Wages for the establishment
1-4 Employees			
	All Occupations (1,798,937 observations)		
	All Controls	0.0041	0.0097
		(0.0004)	(0.0002)
	Lowest-paid quintile of occupations (189,669 observations)		
	All Controls	-0.0034	-0.0125
		(0.0013)	(0.0005)
5-9 Employees			
	All Occupations (3,382,494 observations)		
	All Controls	0.0009	0.0101
		(0.0002)	(0.0001)
	Lowest-paid quintile of occupations (401,487 observations)		
	All Controls	-0.0008	-0.0076
		(0.0008)	(0.0004)
10-19 Employees			
	All Occupations (5,381,267 observations)		
	All Controls	0.0080	0.0008
		(0.0002)	(0.0001)
	Lowest-paid quintile of occupations (605,146 observations)		
	All Controls	0.0114	-0.0153
		(0.0006)	(0.0003)

Occupational Homogeneity Variable	Herfindahl of occupational homogeneity	Predicted Variance of Wages
20-49 Employees		
All Occupations (9,048,401 observations)		
All Controls	0.0094	-0.0007
	(0.0001)	(0.0001)
Lowest-paid quintile of occupations (950,061 observations)		
All Controls	0.0135	-0.0151
	(0.0004)	(0.0002)
50-99 Employees		
All Occupations (7,466,896 observations)		
All Controls	0.0070	0.0030
	(0.0001)	(0.0001)
Lowest-paid quintile of occupations (795,783 observations)		
All Controls	0.0166	-0.0117
	(0.0004)	(0.0002)
100-249 Employees		
All Occupations (8,880,411 observations)		
All Controls	-0.0037	0.0029
	(0.0001)	(0.0000)
Lowest-paid quintile of occupations (888,938 observations)		
All Controls	-0.0071	-0.0159
	(0.0004)	(0.0001)
250-499 Employees		
All Occupations (5,262,100 observations)		
All Controls	-0.0031	0.0075
	(0.0002)	(0.0001)
Lowest-paid quintile of occupations (430,509 observations)		
All Controls	-0.0022	-0.0094
	(0.0006)	(0.0002)
500+ Employees		
All Occupations (5,388,888 observations)		
All Controls	0.0129	0.0100
	(0.0001)	(0.0001)
Lowest-paid quintile of occupations (250,452 observations)		
All Controls	0.0034	0.0019
	(0.0006)	(0.0002)

Note: These are coefficients α from regressions of the form $Occupational\ Homogeneity_{ijt} = \alpha Survey\ Date_t + \beta I(May\ Survey_t) + \gamma X_{ijt} + \varepsilon_{ijt}$, where the Survey Date is measured in decades since X_{ijt} includes occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, establishment size class, continuous establishment size, and state. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Standard errors in parentheses Regressions at the establishment-occupation-wage interval level, weighted by employment.

Table B6: Regressions of log wages on measures of Occupational Homogeneity, by Establishment size group

Wage regressions for quintiles of unimputed OES private-sector data from Nov 2002 - Nov 2016
Occupations are ranked by average wage and classified into quintiles as shown in Appendix A

Occupational Homogeneity Variable	Herfindahl of occupational homogeneity for the establishment	Predicted Variance of Wages for establishment
1-4 Employees		
All Occupations (1,798,937 observations)		
All Controls	-0.064	0.032
	(0.002)	(0.004)
Lowest-paid quintile of occupations (189,669 observations)		
All Controls	-0.095	0.115
	(0.003)	(0.008)
5-9 Employees		
All Occupations (3,382,494 observations)		
All Controls	-0.101	0.071
	(0.001)	(0.002)
Lowest-paid quintile of occupations (401,487 observations)		
All Controls	-0.123	0.119
	(0.002)	(0.004)
10-19 Employees		
All Occupations (5,381,267 observations)		
All Controls	-0.113	0.132
	(0.001)	(0.002)
Lowest-paid quintile of occupations (605,146 observations)		
All Controls	-0.125	0.156
	(0.002)	(0.004)
20-49 Employees		
All Occupations (9,048,401 observations)		
All Controls	-0.107	0.209
	(0.001)	(0.002)
Lowest-paid quintile of occupations (950,061 observations)		
All Controls	-0.100	0.208
	(0.001)	(0.003)
50-99 Employees		
All Occupations (7,466,896 observations)		
All Controls	-0.102	0.327
	(0.001)	(0.002)
Lowest-paid quintile of occupations (795,783 observations)		
All Controls	-0.090	0.315
	(0.002)	(0.005)

Occupational Homogeneity Variable	Herfindahl of occupational homogeneity for the establishment	Predicted Variance of Wages for the establishment
100-249 Employees		
All Occupations (8,880,411 observations)		
All Controls	-0.123	0.406
	(0.001)	(0.002)
Lowest-paid quintile of occupations (888,938 observations)		
All Controls	-0.134	0.365
	(0.002)	(0.005)
250-499 Employees		
All Occupations (5,262,100 observations)		
All Controls	-0.128	0.433
	(0.001)	(0.002)
Lowest-paid quintile of occupations (430,509 observations)		
All Controls	-0.131	0.306
	(0.003)	(0.008)
500+ Employees		
All Occupations (5,388,888 observations)		
All Controls	-0.114	0.345
	(0.001)	(0.003)
Lowest-paid quintile of occupations (250,452 observations)		
All Controls	-0.093	0.242
	(0.004)	(0.011)

Notes: These are coefficients α from regressions of the form $\ln(wage_{ijt}) = \alpha Occupational\ Homogeneity_{jt} + \beta OccConcen_{jt}Date_t + \gamma X_{ijt} + \varepsilon_{ijt}$, where Date is measured in decades since November 1, 2002, X includes survey date fixed effects, occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, state fixed effects, and establishment size (using fixed effects for establishment size classes as well as continuous establishment size). All coefficients are significant at $p < 0.001$. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Standard errors in parentheses.

B4: Heterogeneity by industrial sector

Dividing establishments into industrial sectors, establishments in the Construction (23), Administrative and Support and Waste Management (56), Accommodation and Food Services (72), and Other Services (except Public Administration) (81) have high levels of occupational homogeneity, by all measures. Management of Companies and Enterprises (55) has particularly low levels of occupational homogeneity, by all measures, while Utilities (22) has particularly low occupational homogeneity by the Herfindahl-Hirschman measure, and private-sector Educational Services (61) has particularly low occupational homogeneity by the total predicated variance of wages based on the occupational distribution.

These sectors differ greatly in the fraction of their employment in occupations at different places in the wage distribution, but some of the same sectors stand out for having low or high levels of occupational homogeneity among employers of workers in occupations in the lowest-paid quintile of the labor force. Examining occupational homogeneity levels by sector for workers in these low-paid occupations, again, Administrative and Support and Waste Management (56) and Other Services (except Public Administration) (81) have particularly high levels of occupational homogeneity, by all measures and Management of Companies and Enterprises (55) has particularly low levels of occupational homogeneity, while private-sector Educational Services (61) has particularly low occupational homogeneity by the total predicated variance of wages based on the occupational distribution. However, other sectors with particularly high or low levels of occupational homogeneity are different for workers in the bottom-paid quintile of occupations. For these workers, establishments in the Educational Services (61) sector also have particularly low levels of occupational homogeneity by all measures. Establishments in the Finance and Insurance Sector (72) have a high level of occupational homogeneity by the Herfindahl measure, while the Accommodation and Food Services (72) and Other Services (42) sectors have particularly high levels of occupational homogeneity by both predicted variance measures.

Trends over time in occupational homogeneity vary greatly by sector. Across all occupations and all measures of occupational homogeneity, after controlling for observable characteristics, occupational homogeneity is increasing within the Accommodation and Food Services (72) and Transportation and Warehousing (48) sectors. It is decreasing within the Construction (23), (private-sector) Educational Services (61), Manufacturing (31-33), Finance and Insurance (52), Real Estate and Rental and Leasing (53), and Other Services (81) sectors, with trends in the remaining sectors that vary in direction by measure of occupational homogeneity. For the lowest-paid quintile of occupations, after controlling for observable characteristics, occupational homogeneity is increasing by all measures within the Construction (23), Administrative and Support and Waste Management (56), Arts, Entertainment, and Recreation (71), and Accommodation and Food Services (72) sectors, while it is decreasing by all measures only within the Management of Companies and Enterprises (55) and (private-sector) Educational Services (61) sectors.

The relationship between occupational homogeneity and wages also varies tremendously by sector, even after controlling for the occupations employed within each sector. After controlling for occupations, detailed industries, state, and establishment size, by every measure, greater occupational homogeneity is associated with lower wages within the Construction (23), Retail Trade (44), Transportation and Warehousing (48), Real Estate and Rental and Leasing (53), Professional, Scientific, and Tech Services (54), Management of Companies and Enterprises (55), Administrative and Support and Waste Management (56), Health Care and Social Assistance (62), Arts, Entertainment, and Recreation (71), Accommodation and Food Services (72), and Other Services (81) sectors. These relationships are particularly strong in the Transportation and Warehousing (48) sector. Other sectors have overall relationships between occupational homogeneity and wages that vary by measure of occupational homogeneity. For bottom-quintile workers, after controlling for observable characteristics, greater occupational homogeneity—by every measure—is associated with lower wages within all of the above sectors, as well as the Wholesale Trade (42) and (private-sector) Educational Services (61) sectors. For typically low-wage workers, the relationship between occupational homogeneity and wages is particularly strong across all measures of occupational homogeneity within the Utilities (22) and Arts, Entertainment, and Recreation (71) sectors.

Table B7: Change in mean values of Occupational Homogeneity over time, by sector

Trend regressions for quintiles of unimputed OES private-sector data from Nov 2002-Nov 2016
Occupations are ranked by average wage and classified into quintiles as shown in Appendix A

Occupational Homogeneity Variable	Herfindahl of occupational homogeneity for the establishment	Predicted Variance of Wages for the establishment
Establishments in the Construction Sector		
All Occupations (2,715,765 observations) and all available controls		
All Controls	-0.0069	0.0091
	(0.0003)	(0.0001)
Lowest-paid quintile of occupations (18,771 observations) and all available controls		
All Controls	0.0168	-0.0044
	(0.0032)	(0.0017)
Establishments in the Manufacturing Sector		
All Occupations (7,617,094 observations) and all available controls		
All Controls	-0.0049	0.0157
	(0.0001)	(0.0001)
Lowest-paid quintile of occupations (120,306 observations) and all available controls		
All Controls	-0.0066	-0.0060
	(0.0011)	(0.0006)
Establishments in the Wholesale Trade Sector		
All Occupations (3,121,979 observations) and all available controls		
All Controls	0.0020	0.0176
	(0.0002)	(0.0001)
Lowest-paid quintile of occupations (111,547 observations) and all available controls		
All Controls	0.0031	0.0033
	(0.0013)	(0.0007)
Establishments in the Retail Trade Sector		
All Occupations (6,033,799 observations) and all available controls		
All Controls	0.0061	-0.0166
	(0.0002)	(0.0001)
Lowest-paid quintile of occupations (1,631,641 observations) and all available controls		
All Controls	0.0008	-0.0224
	(0.0003)	(0.0001)
Establishments in the Transportation and Warehousing Sector		
All Occupations (1,372,852 observations) and all available controls		
All Controls	0.0207	-0.0125
	(0.0004)	(0.0002)
Lowest-paid quintile of occupations (23,087 observations) and all available controls		
All Controls	0.0058	0.0021
	(0.0030)	(0.0014)
Establishments in the Information Sector		
All Occupations (1,509,864 observations) and all available controls		
All Controls	0.0050	0.0204
	(0.0004)	(0.0001)
Lowest-paid quintile of occupations (36,871 observations) and all available controls		
All Controls	0.0296	-0.0009
	(0.0023)	(0.0012)
Establishments in the Finance and Insurance Sector		
All Occupations (2,466,725 observations) and all available controls		
All Controls	-0.0053	0.0218
	(0.0002)	(0.0001)
Lowest-paid quintile of occupations (18,020 observations) and all available controls		
All Controls	0.0020	-0.0079
	(0.0037)	(0.0021)
Establishments in the Real Estate and Rental and Leasing Sector		
All Occupations (898,150 observations) and all available controls		
All Controls	-0.0039	0.0137
	(0.0005)	(0.0002)
Lowest-paid quintile of occupations (116,261 observations) and all available controls		
All Controls	-0.0103	-0.0044
	(0.0015)	(0.0007)

Table B7: Change in mean values of Occupational Homogeneity over time, by sector, cont.

Occupational Homogeneity Variable	Herfindahl of occupational homogeneity for the establishment	Predicted Variance of Wages for the establishment
Establishments in the Professional, Scientific, and Tech Servics Sector		
All Occupations (3,501,795 observations) and all available controls		
All Controls	0.0128	0.0160
	(0.0002)	(0.0001)
Lowest-paid quintile of occupations (31,438 observations) and all available controls		
All Controls	0.0037	0.0050
	(0.0023)	(0.0014)
Establishments in the Management of Companies and Enterprises Sector		
All Occupations (1,276,951 observations) and all available controls		
All Controls	0.0006	0.0186
	(0.0003)	(0.0001)
Lowest-paid quintile of occupations (14,398 observations) and all available controls		
All Controls	-0.0900	0.0485
	(0.0047)	(0.0025)
Establishments in the Administrative and Support and Waste Mgmnt Sector		
All Occupations (2,304,399 observations) and all available controls		
All Controls	-0.0005	0.0047
	(0.0003)	(0.0001)
Lowest-paid quintile of occupations (129,160 observations) and all available controls		
All Controls	0.0107	-0.0038
	(0.0012)	(0.0005)
Establishments in the Educational Services Sector		
All Occupations (1,298,489 observations) and all available controls		
All Controls	-0.0045	0.0260
	(0.0003)	(0.0001)
Lowest-paid quintile of occupations (65,498 observations) and all available controls		
All Controls	-0.0263	0.0221
	(0.0013)	(0.0007)
Establishments in the Health Care and Social Assistance Sector		
All Occupations (6,809,331 observations) and all available controls		
All Controls	0.0125	0.0090
	(0.0001)	(0.0001)
Lowest-paid quintile of occupations (614,503 observations) and all available controls		
All Controls	0.0046	-0.0061
	(0.0005)	(0.0002)
Establishments in the Arts, Entertainment, and Recreation Sector		
All Occupations (1,062,174 observations) and all available controls		
All Controls	0.0088	0.0049
	(0.0004)	(0.0002)
Lowest-paid quintile of occupations (318,623 observations) and all available controls		
All Controls	0.0142	-0.0047
	(0.0007)	(0.0004)
Establishments in the Accommodation and Food Services Sector		
All Occupations (2,078,374 observations) and all available controls		
All Controls	0.0081	-0.0090
	(0.0003)	(0.0001)
Lowest-paid quintile of occupations (1,050,145 observations) and all available controls		
All Controls	0.0097	-0.0099
	(0.0004)	(0.0002)
Establishments in the Other Services (except Public Admin) Sector		
All Occupations (1,876,169 observations) and all available controls		
All Controls	-0.0041	0.0078
	(0.0003)	(0.0001)
Lowest-paid quintile of occupations (206,167 observations) and all available controls		
All Controls	-0.0094	-0.0010
	(0.0010)	(0.0005)

Table B8: Regressions of log wages on measures of Occupational Homogeneity, by sector

Wage regressions for quintiles of unimputed OES private-sector data from Nov 2002 - Nov 2016

Occupations are ranked by average wage and classified into quintiles as shown in Appendix A

Occupational Homogeneity Variable	Herfindahl of occupational homogeneity for the establishment	Predicted Variance of Wages for the establishment
Establishments in the Construction Sector		
All Occupations (2,715,765 observations) and all available controls		
All Controls	-0.008	1.035
	(0.011)	(0.042)
Lowest-paid quintile of occupations (18,771 observations) anc all available controls		
All Controls	-0.274	1.932
	(0.161)	(0.316)
Establishments in the Manufacturing Sector		
All Occupations (7,617,094 observations) and all available controls		
All Controls	0.001	0.118
	(0.007)	(0.018)
Lowest-paid quintile of occupations (120,306 observations) and all available controls		
All Controls	-0.276	0.034
	(0.044)	(0.096)
Establishments in the Wholesale Trade Sector		
All Occupations (3,121,979 observations) and all available controls		
All Controls	0.064	0.219
	(0.012)	(0.034)
Lowest-paid quintile of occupations (111,547 observations) and all available controls		
All Controls	-0.558	0.020
	(0.054)	(0.120)
Establishments in the Retail Trade Sector		
All Occupations (6,033,799 observations) and all available controls		
All Controls	-0.463	0.092
	(0.007)	(0.018)
Lowest-paid quintile of occupations (1,631,641 observations) and all available controls		
All Controls	-0.609	0.205
	(0.011)	(0.033)
Establishments in the Transportation and Warehousing Sector		
All Occupations (1,372,852 observations) and all available controls		
All Controls	-0.350	1.465
	(0.016)	(0.028)
Lowest-paid quintile of occupations (23,087 observations) anc all available controls		
All Controls	-0.049	1.121
	(0.121)	(0.249)
Establishments in the Information Sector		
All Occupations (1,509,864 observations) and all available controls		
All Controls	-0.108	-0.362
	(0.017)	(0.045)
Lowest-paid quintile of occupations (36,871 observations) anc all available controls		
All Controls	-0.425	0.531
	(0.071)	(0.145)
Establishments in the Finance and Insurance Sector		
All Occupations (2,466,725 observations) and all available controls		
All Controls	0.142	0.489
	(0.014)	(0.040)
Lowest-paid quintile of occupations (18,020 observations) anc all available controls		
All Controls	0.075	0.033
	(0.103)	(0.237)
Establishments in the Real Estate and Rental and Leasing Sector		
All Occupations (898,150 observations) and all available controls		
All Controls	-0.282	0.979
	(0.019)	(0.047)
Lowest-paid quintile of occupations (116,261 observations) and all available controls		
All Controls	-0.466	0.567
	(0.039)	(0.095)

Table B8: Regressions of log wages on measures of Occ Homogeneity, by sector continued

Occupational Homogeneity Variable	Herfindahl of occupational homogeneity for the establishment	Predicted Variance of Wages for the establishment
Establishments in the Professional, Scientific, and Tech Servics Sector		
All Occupations (3,501,795 observations) and all available controls		
All Controls	-0.016	0.781
	(0.011)	(0.023)
Lowest-paid quintile of occupations (31,438 observations) and all available controls		
All Controls	-0.994	0.833
	(0.095)	(0.185)
Establishments in the Management of Companies and Enterprises Sector		
All Occupations (1,276,951 observations) and all available controls		
All Controls	-0.603	0.453
	(0.027)	(0.057)
Lowest-paid quintile of occupations (14,398 observations) and all available controls		
All Controls	-0.091	2.595
	(0.127)	(0.254)
Establishments in the Administrative and Support and Waste Mgmnt Sector		
All Occupations (2,304,399 observations) and all available controls		
All Controls	-0.156	0.720
	(0.009)	(0.026)
Lowest-paid quintile of occupations (129,160 observations) and all available controls		
All Controls	-0.357	0.859
	(0.035)	(0.101)
Establishments in the Educational Services Sector		
All Occupations (1,298,489 observations) and all available controls		
All Controls	0.127	-0.188
	(0.020)	(0.058)
Lowest-paid quintile of occupations (65,498 observations) and all available controls		
All Controls	-0.379	0.612
	(0.059)	(0.136)
Establishments in the Health Care and Social Assistance Sector		
All Occupations (6,809,331 observations) and all available controls		
All Controls	-0.127	0.231
	(0.006)	(0.013)
Lowest-paid quintile of occupations (614,503 observations) and all available controls		
All Controls	-0.222	1.072
	(0.011)	(0.029)
Establishments in the Arts, Entertainment, and Recreation Sector		
All Occupations (1,062,174 observations) and all available controls		
All Controls	-0.670	0.495
	(0.018)	(0.039)
Lowest-paid quintile of occupations (318,623 observations) and all available controls		
All Controls	-0.787	0.999
	(0.027)	(0.069)
Establishments in the Accommodation and Food Services Sector		
All Occupations (2,078,374 observations) and all available controls		
All Controls	-0.075	0.698
	(0.009)	(0.025)
Lowest-paid quintile of occupations (1,050,145 observations) and all available controls		
All Controls	0.023	0.348
	(0.011)	(0.034)
Establishments in the Other Services (except Public Admin) Sector		
All Occupations (1,876,169 observations) and all available controls		
All Controls	-0.145	0.144
	(0.011)	(0.032)
Lowest-paid quintile of occupations (206,167 observations) and all available controls		
All Controls	-0.161	0.558
	(0.028)	(0.075)

B5: Heterogeneity by Employer tax Identification Number (EIN) size

Song et al find very different patterns of inequality growth from 1978 to 2013 for very large firms—those with Employer tax Identification Numbers (EINs) with 10,000 or more employees—than for smaller firms. They find that for smaller firms, nearly all inequality growth is between firms, explained by greater sorting of workers with higher worker fixed effects to firms with higher firm fixed effects, while very larger firms see nearly half of inequality growth happening within firms, with falling wages for their lowest-paid workers and rising wages for their highest-paid workers. The focus of this analysis is the establishment (except in section V) because establishments are the sampling units of the OES, and the OES sampling design often includes some but not all establishments of multi-establishment companies, particularly when there are establishments in geographic areas with fewer establishments available to sample. However, the OES microdata can be linked with the EIN numbers that these establishments submit to the unemployment insurance system, as compiled by the Quarterly Census of Employment and Wages (QCEW), and the full employment level of each EIN can be calculated in each time period using the QCEW data. As discussed in Handwerker and Mason (2013), very large firms may use multiple EINs in the unemployment insurance system, and there is no way to link together all of the establishments in these data for very large firms without a tremendous amount of manual review. It is straightforward to group establishments into those that are part of very large EINs (those with 10,000 or more employees) and those that are not part of these large EINs, but the reader should be aware that many establishments not part of large EINs are nonetheless part of large firms that use smaller EINs in their quarterly reports to the unemployment insurance system.[14]

The reader should also be aware that very large establishments are included in the OES sample with certainty every 6 panels,[15] and the reweighting used to break apart the 6-panel groups of OES waves used for official OES publications into panel-specific microdata results in big swings of estimates (of any variable of interest) from one panel to the next for very large employers. This makes it difficult to measure trends in any variable for these employers. Nonetheless, the OES data show that workers in the bottom quintile of occupations were paid more in huge firms than in smaller firms during earlier waves of data collection, but this difference disappeared around November 2013. This is consistent with the finding of Song et. al. that workers with low values of worker fixed effects in very large firms have seen declining wages over time. It is not exactly comparable to Song et. al. because those authors use repeated observations of workers over time to estimate worker fixed effects, an estimation not possible with the OES data. However, there is likely a great deal of overlap between workers in typically-low-wage occupations and workers with low fixed effects.

[14] Song et al estimate that 23% of workers are in these very large firms; in the OES data, only 16% of workers are in such large firms. Various BLS projects have attempted to find all the EINs associated with particular sets of firms in particular time periods. Using all of the available links from these projects to group EINs together, it is possible to consider 20% of workers as associated with firms employing 10,000 or more; results using these larger groupings are little changed from the EIN-level results discussed in this section.

[15] Where the sample design allows it, the OES program also makes an effort to sample establishments of the same employer in the same wave of the sample. Thus, the establishments of individual large employers are likely to appear in the OES data every six panels.

Overall, the establishments of very large employers have lower Herfindahl-Hirschman values (.36, weighted by employment) of occupational homogeneity than the establishments of smaller employers (.42), and this is especially true for workers in the bottom quintile of occupations (.44 for very large employers and .50 for smaller employers). However, the predicted variance of wages based on occupational distributions is greater for the establishments of smaller employers (.27) than for very large employers (.26), although this reverses for workers in the bottom quintile (.20 for smaller employers and .21 for very large employers). The establishments of very large employers are much less likely than the establishments of smaller employers to have no workers in the top quintile of occupations (13% of the employees of very large employers have no top-quintile co-workers, compared with 25% of the employees of smaller establishments), and this is particularly true for workers in the bottom quintile (25% of the bottom-quintile employees of very large employers have no top-quintile co-workers, compared with 46% of the bottom-quintile employees of smaller employers). Relative to smaller employers, very large employers have establishments that are more diverse in the groups of occupations they employ, but are more homogeneous in the wage range of the occupations they employ—but the very large employers of workers of the bottom quintile of occupations workers are a little less homogeneous in the wage range of occupations they employ than smaller employers of these low-paid occupations.

Trends in occupational homogeneity differ between very large employers and smaller employers for smaller employers, for each measure and for all occupations as well as for the lowest-paid quintile of occupations, with and without controlling for other observable variables. By all measures, equation (3) shows small but significant increases in occupational homogeneity within huge employers for every measure of occupational homogeneity. This is the opposite of the trend found for small employers for the predicted variance of wages (and its between-establishment component) measures.

Overall, occupational homogeneity matters much more for wages within huge employers, by every measure of occupational homogeneity. However, this difference reverses once controls for establishment characteristics and occupations are added to equation (4), or if the occupations in the lowest-paid quintile are examined separately.

Table B9: Change in mean values of Occupational Homogeneity over time, by EIN size

Trend regressions for quintiles of unimputed OES private-sector data from Nov 2002-Nov 2016

Occupations are ranked by average wage and classified into quintiles as shown in Appendix A

Occupational Homogeneity Variable	Herfindahl of occupational homogeneity for the establishment	Predicted Variance of Wages for the establishment
Establishments associated with huge (10,000+) employment EINs		
All Occupations (6,028,858 observations)		
Raw trend	0.0116	-0.0032
	(0.00021)	(0.00010)
All Controls	0.0109	-0.0062
	(0.00015)	(0.00006)
Lowest-paid quintile of occupations (928,376 observations)		
Raw trend	0.0168	-0.0236
	(0.00055)	(0.00019)
All Controls	0.0127	-0.0211
	(0.00039)	(0.00014)
Establishments associated with smaller EINs		
All Occupations (40,580,536 observations)		
Raw trend	0.0064	0.0079
	(0.00010)	(0.00004)
All Controls	0.0071	0.0065
	(0.00008)	(0.00003)
Lowest-paid quintile of occupations (3,583,669 observations)		
Raw trend	0.0169	-0.0152
	(0.00034)	(0.00013)
All Controls	0.0085	-0.0108
	(0.00025)	(0.00010)

Note: These are coefficients α from regressions of the form $Occupational\ Homogeneity_{ijt} = \alpha Survey\ Date_t + \beta I(May\ Survey_t) + \gamma X_{ijt} + \varepsilon_{ijt}$, where the Survey Date is measured in decades since X_{ijt} includes occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, establishment size class, continuous establishment size, and state. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Standard errors in parentheses Regressions at the establishment-occupation-wage interval level, weighted by employment.

Table B10: Regressions of log wages on measures of Occupational Homogeneity, by EIN size

Wage regressions for quintiles of unimputed OES private-sector data from Nov 2002 - Nov 2016
Occupations are ranked by average wage and classified into quintiles as shown in Appendix A

Occupational Homogeneity Variable	Herfindahl of occupational homogeneity for the establishment	Predicted Variance of Wages for the establishment
Establishments associated with huge (10,000+) employment EINs		
All Occupations (6,028,858 observations)		
With only date fixed effects	-0.737	2.547
	(0.002)	(0.005)
All Controls	-0.067	0.228
	(0.001)	(0.003)
Lowest-paid quintile of occupations (928,376 observations)		
With only date fixed effects	-0.243	0.817
	(0.003)	(0.008)
All Controls	-0.102	0.276
	(0.003)	(0.009)
Establishments associated with smaller EINs		
All Occupations (40,580,536 observations)		
With only date fixed effects	-0.472	2.252
	(0.001)	(0.002)
All Controls	-0.108	0.242
	(0.000)	(0.001)
Lowest-paid quintile of occupations (3,583,669 observations)		
With only date fixed effects	-0.219	0.876
	(0.001)	(0.003)
All Controls	-0.139	0.346
	(0.001)	(0.003)

Notes: These are coefficients α from regressions of the form $\ln(wage_{ijt}) = \alpha Occupational\ Homogeneity_{jt} + \beta OccConcen_{jt} Date_t + \gamma X_{ijt} + \varepsilon_{ijt}$, where Date is measured in decades since November 1, 2002, X includes survey date fixed effects, occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, state fixed effects, and establishment size (using fixed effects for establishment size classes as well as continuous establishment size). All coefficients are significant at $p < 0.001$. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Standard errors in parentheses.

Appendix C: Occupational Homogeneity: Establishment or Firm Measures?

Song, Price, Guvenen, Bloom, and von Wachter (2016) argue that the unit of importance for wage inequality should be the firm and not the establishment. In thinking about occupational homogeneity, some of the reasons for employers to outsource work to other establishments are also reasons to outsource work to other employers entirely. It may be more efficient for even multi-establishment employers to specialize in particular areas of work. Regulatory incentives for multi-establishment employers to specialize in employing workers in a particular part of the wage distribution are less clear. ERISA laws define employers as "controlled groups of corporations" and "entities under common control" in requiring common levels of pension and welfare benefits among most employees in exchange for favorable tax treatment (Perun 2010), and the Affordable Care Act of 2010 extended these provisions by requiring common levels of health care benefits among most employees of businesses with a common owner. However, as Perun notes, "Employers often invent new organizational structures and worker classifications designed to limit participation to favored employees… Regulatory authorities in turn develop complicated rules and regulations designed to prevent this."

This paper focuses on measures of occupational homogeneity at the establishment level because establishments are the sampling units of the OES, and the OES sampling design often includes some but not all establishments of multi-establishment companies, particularly when there are establishments in geographic areas with fewer establishments available to sample. However, the OES microdata can be linked with the EIN (tax-ID) numbers that these establishments submit to the unemployment insurance system, as compiled by the Quarterly Census of Employment and Wages. As discussed extensively in Handwerker and Mason (2013), very large firms may use multiple EINs in the unemployment insurance system, and there is no way to link together all of the establishments in these data for very large firms without a tremendous amount of manual review. Thus, while it is straightforward to recalculate measures of occupational homogeneity at the EIN level and repeat the analyses above, such EIN-level measures are not true firm-level measures.

Using EIN-level measures of occupational homogeneity instead of establishment-level measures has remarkably little impact on any of the main results in this paper.[16] Trends in EIN-level measures of occupational homogeneity over time are very similar to those for establishment-level measures in Table 2, shown in Table C1. However, the increased occupational homogeneity of workers in the bottom quintile is significantly larger when measured with establishment-level measures of occupational homogeneity than when using the EIN-level equivalents. The relationship between EIN-level measures of occupational homogeneity and wages, shown in Table C2, is very similar to that shown for establishment-level measures in Table 4.

Reweighting the November 2014/May 2015 data to the November 2002/May 2003 distribution of EIN-level measures of occupational homogeneity also yields very similar results to those shown in Table 5, as shown in Table C3.

[16] Note that 92% of the EINs in the OES data (containing 64% of weighted employment) contain data from only one establishment per data panel. However, trends in occupational homogeneity are very similar for employers with 10 or more establishments per data panel.

Appendix Table C1: Change in mean values of EIN-level Occupational Homogeneity over time

Trend regressions for quintiles of unimputed OES private-sector data from Nov 2002-Nov 2016

Occupations are ranked by average wage and classified into quintiles as shown in Appendix A

Occupational Homogeneity Variable	Herfindahl of occupational homogeneity for the EIN		Predicted Variance of Wages for the EIN	
All Occupations (45,689,947 observations)				
Raw trend	0.0002	*	0.0075	***
	(0.00009)		(0.00004)	
All Controls	0.0009	***	0.0065	***
	(0.00006)		(0.00003)	
Lowest-paid quintile of occupations (4,394,954 observations)				
Raw trend	0.0127	***	-0.0150	***
	(0.00030)		(0.00011)	
All Controls	0.0026	***	-0.0104	***
	(0.00021)		(0.00009)	
Second quintile of occuaptions (6,580,804 observations)				
Raw trend	0.0101	***	0.0001	
	(0.00024)		(0.00009)	
All Controls	0.0032	***	0.0044	***
	(0.00016)		(0.00007)	
Middle quintile of occupations (9,564,315 observations)				
Raw trend	0.0014	***	0.0104	***
	(0.00017)		(0.00008)	
All Controls	-0.0009	***	0.0116	***
	(0.00013)		(0.00006)	
Fourth quintile of occupations (11,653,657 observations)				
Raw trend	-0.0161	***	0.0158	***
	(0.00018)		(0.00006)	
All Controls	-0.0032	***	0.0117	***
	(0.00013)		(0.00004)	
Highest-paid quintile of occupations (13,496,217 observations)				
Raw trend	-0.0009	***	0.0202	***
	(0.00014)		(0.00006)	
All Controls	0.0027	***	0.0174	***
	(0.00010)		(0.00004)	

Note: These are coefficients α from regressions of the form $Occupational\ Homogeneity_{ijt} = \alpha Survey\ Date_t + \beta I(May\ Survey_t) + \gamma X_{ijt} + \varepsilon_{ijt}$, where the Survey Date is measured in decades since X_{ijt} includes occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, establishment size class, continuous establishment size, and state. "*" indicates p<0.05; "**" indicates p<0.01; and "***" indicates p<0.001. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Standard errors in parentheses. Regressions at the establishment-occupation-wage interval level, weighted by employment.

Appendix Table C2: Regressions of log wages on EIN-level measures of Occupational Homogeneity

Wage regressions for quintiles of unimputed OES private-sector data from Nov 2002 - Nov 2016

Occupations are ranked by average wage and classified into quintiles as shown in Appendix A

Occupational Homogeneity Variable	Herfindahl of occupational homogeneity for the EIN	Predicted Variance of Wages for the EIN
All Occupations (45,689,947 observations)		
With only date fixed effects	-0.495	2.315
	(0.001)	(0.002)
All Controls	-0.133	0.296
	(0.000)	(0.001)
Lowest-paid quintile of occupations (4,394,954 observations)		
With only date fixed effects	-0.229	0.881
	(0.001)	(0.003)
All Controls	-0.141	0.369
	(0.001)	(0.003)
Second quintile of occupations (6,580,804 observations)		
With only date fixed effects	-0.286	0.786
	(0.001)	(0.003)
All Controls	-0.202	0.451
	(0.001)	(0.002)
Middle quintile of occupations (9,564,315 observations)		
With only date fixed effects	-0.221	0.612
	(0.001)	(0.002)
All Controls	-0.167	0.370
	(0.001)	(0.002)
Fourth quintile of occupations (11,653,657 observations)		
With only date fixed effects	-0.295	0.755
	(0.001)	(0.003)
All Controls	-0.108	0.199
	(0.001)	(0.003)
Highest-paid quintile of occupations (13,496,217 observations)		
With only date fixed effects	-0.195	0.526
	(0.001)	(0.003)
All Controls	-0.009	-0.093
	(0.001)	(0.003)

Notes: These are coefficients α from regressions of the form $\ln(wage_{ijt}) = \alpha Occupational\ Homogeneity_{jt} + \beta OccConcen_{jt} Date_t + \gamma X_{ijt} + \varepsilon_{ijt}$, where Date is measured in decades since November 1, 2002, X includes survey date fixed effects, occupation fixed effects at the 6-digit SOC level, industry fixed effects at the 4-digit NAICS level, state fixed effects, and establishment size (using fixed effects for establishment size classes as well as continuous establishment size). All coefficients are significant at $p < 0.001$. Regressions are at the establishment-occupation-wage interval level, weighted by employment. Standard errors in parentheses.

Appendix Table C3: Decomposition of Changes in real ln(wage) variance from 2002-2003 to 2014-2015, using EIN-level measures of Occupational Homogeneity and employer size

real log wage variance	Coeff.	Percent	Std. Err.	z	P>|z|	[95% Conf. Interval]	
Overall Variance							
Late period (2014-2015)	0.3999		0.0013	314.4	0	0.3974	0.4024
counterfactual variance	0.3691		0.0009	391	0	0.3672	0.3709
Early period (2002-2003)	0.3754		0.0016	232	0	0.3722	0.3786
Total change	0.0245	100%	0.0021	11.89	0	0.0204	0.0285
of which explained (by compositional change)	0.0308	126%	0.0006	54.87	0	0.0297	0.0319
of which unexplained (wage structure change)	-0.0063	-26%	0.0019	-3.32	0.001	-0.0101	-0.0026
Explained (compositional effect)							
Total	0.0308	100%	0.0006	54.87	0	0.0297	0.0319
Pure_explained	0.0295	96%	0.0006	46.92	0	0.0282	0.0307
Specification_error	0.0013	4%	0.0001	10.01	0	0.0011	0.0016
Components of the pure explained effect							
industrial sector	0.0000	0%	0.0002	0.16	0.87	-0.0003	0.0004
geography (Census division)	-0.0001	0%	0.0001	-1.59	0.111	-0.0003	0.0000
size	0.0013	4%	0.0002	7.45	0	0.0009	0.0016
quintiles of occupation	0.0155	50%	0.0006	27	0	0.0144	0.0166
Establishment Herfindahls	0.0024	8%	0.0005	5.13	0	0.0015	0.0033
Establishment predicted var(ln wages)	0.0105	34%	0.0007	14.79	0	0.0091	0.0118
Unexplained (wage structure changes)							
Total	-0.0063	100%	0.0019	-3.32	0.001	-0.0101	-0.0026
Reweight_error	-0.0001	2%	0.0008	-0.12	0.901	-0.0017	0.0015
Pure_Unexplained	-0.0062	98%	0.0017	-3.61	0	-0.0096	-0.0028

Notes: These are the results of Rios-Avila's implementation of Fortin, Firpo, and Lemiueux's Recentered Influence Function Decomposition. Industry here is grouped into 19 supersectors and geography into 7 Census divisions. EIN size is measured in 8 categories (1-4 employees, 5-9, 10-19, 20-49, 50-99, 100-249, 250-499, and 500+). EIN-level Herfindahl-Hirschman indices of occupations are measured with quartiles of the distribution, interacted with occupational quintiles. The predicted variances of ln(wage) for EINs is divided into quartiles, and also interacted with occupational quintiles, with an additional dummy variable for low-wage occupations in EINs of less than 100 workers that are in the bottom half of the predicted variance distribution. Standard errors presented here have not been bootstrapped.

Appendix D: Alternative RIF Decompositions of the change in real ln(wage) variance over time

This Appendix shows two variations on Table 5: a decomposition of wage variance change using greater detail in the explanatory variables, and a decomposition using a later end date.

D1: Greater Detail in the Explanatory Variables

The main RIF decomposition, shown in Table 5, uses much coarser measures of industry, occupation, and geography than the regression controls of Tables 2-4. This is to speed the processing time involved in carrying out the RIF decomposition so that the standard errors can be bootstrapped. The results of an alternate RIF decomposition with detailed measures of industry, occupation, and geography, as well as finer categories of occupational homogeneity are shown in Table D1. Here, only 76% of the change in ln(wage) variance can be explained by compositional changes in employer and employee characteristics. However, of the amount that can be explained, changes in occupation and of employer homogeneity (by the predicted variance of wages measure) continue to be by far the most important categories of variables. Because changes in wage structure explain 24% of the change in ln(wage) variance, the decomposition of wage structure is also shown here by category.

Table D1: Decomposition of Changes in real ln(wage) variance from 2002-2003 to 2014-2015, with greater detail used in the explanatory variables

real log wage variance	Coeff.	Percent	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Overall Variance							
Late period (2014-2015)	0.3999		0.0006	644.42	0	0.3987	0.4011
counterfactual variance	0.3812		0.0006	638.16	0	0.3801	0.3824
Early period (2002-2003)	0.3754		0.0009	433.03	0	0.3737	0.3771
Total change	0.0245	100%	0.0011	22.96	0	0.0224	0.0266
of which explained (by compositional change)	0.0186	76%	0.0003	72.55	0	0.0181	0.0191
of which unexplained (wage structure change)	0.0058	24%	0.0011	5.55	0	0.0038	0.0079
Explained (compositional effect)							
Total	0.0186	100%	0.0003	72.55	0	0.0181	0.0191
Pure_explained	0.0175	94%	0.0003	66.15	0	0.0170	0.0180
Specification_error	0.0011	6%	0.0001	15.66	0	0.0010	0.0012
Components of the pure explained effect							
industry (4-digit NAICS)	-0.0007	-3%	0.0001	-5.6	0	-0.0009	-0.0004
geography (state)	-0.0013	-7%	0.0000	-34.3	0	-0.0014	-0.0013
size	0.0020	11%	0.0001	39.24	0	0.0019	0.0021
occupation (minor occupational categories)	0.0086	46%	0.0002	41.6	0	0.0082	0.0090
Establishment Herfindahls	-0.0007	-4%	0.0001	-7.26	0	-0.0009	-0.0005
Establishment predicted var(ln wages)	0.0096	52%	0.0001	81.28	0	0.0094	0.0099
Unexplained (wage structure changes)							
Total	0.0058	100%	0.0011	5.55	0	0.0038	0.0079
Reweight_error	-0.0011	-18%	0.0010	-1.07	0.283	-0.0030	0.0009
Pure_Unexplained	0.0069	118%	0.0005	15.3	0	0.0060	0.0078
Components of the pure unexplained effect							
industry (4-digit NAICS)	-0.2221	-3799%	0.0002	-1199.3	0	-0.2225	-0.2217
geography (state)	0.0732	1253%	0.0001	1272.7	0	0.0731	0.0734
size	0.0358	612%	0.0001	426.22	0	0.0356	0.0359
occupation (minor occupational categories)	0.0237	405%	0.0005	45	0	0.0227	0.0247
Establishment Herfindahls	-0.2351	-4021%	0.0003	-792.55	0	-0.2357	-0.2345
Establishment predicted var(ln wages)	-0.0101	-172%	0.0004	-24.01	0	-0.0109	-0.0092
_cons	0.3415	5841%	.	.	.	.	.

Notes: These are the results of Rios-Avila’s implementation of Fortin, Firpo, and Lemiueux’s Recentered Influence Function Decomposition. Industry here is grouped into 280 time-consistent 4-digit NAICS codes and geography by state. Establishment size is measured in 8 categories (1-4 employees, 5-9, 10-19, 20-49, 50-99, 100-249, 250-499, and 500+). Establishment-level Herfindahl-Hirschman indices of occupations are measured with deciles of the distribution, interacted with occupational quintiles. The predicted variance of ln(wage) for establishments is divided into deciles, and also interacted with occupational quintiles, with an additional dummy variable for low-wage occupations in establishment of less than 100 workers that are in the bottom half of the predicted variance distribution. Standard errors have not been bootstrapped.

D2: A later end date

The latest panel of data used in this paper is data collected in November 2016. For reasons outside the scope of this paper, the overall level of wage variance in the OES microdata was higher than trend in November 2002, and lower than trend in November 2016. Thus, there is not much increase in wage variance to explain between these end dates, and changes in the composition of employers and employees can explain 243% of the change in wage variance (as shown in Table D2). However, the strong role of changes in the composition of the labor force by quintile (employment polarization) and of employer homogeneity remain.

Table D2: Decomposition of Changes in real ln(wage) variance from Nov 2002 & May 2003 to May 2016 & Nov 2016

real log wage variance	Coeff.	Percent	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Overall Variance							
Late period (2016)	0.3877		0.0013	303.81	0	0.3852	0.3902
counterfactual variance	0.3583		0.0010	353.2	0	0.3563	0.3603
Early period (2002-2003)	0.3756		0.0016	233.01	0	0.3725	0.3788
Total change	0.0121	100%	0.0021	5.88	0	0.0081	0.0161
of which explained (by compositional change)	0.0294	243%	0.0005	60.77	0	0.0284	0.0303
of which unexplained (wage structure change)	-0.0173	-143%	0.0019	-8.92	0	-0.0211	-0.0135
Explained (compositional effect)							
Total	0.0294	100%	0.0005	60.77	0	0.0284	0.0303
Pure_explained	0.0291	99%	0.0005	57.73	0	0.0281	0.0301
Specification_error	0.0003	1%	0.0001	3.15	0.002	0.0001	0.0004
Components of the pure explained effect							
industrial sector	-0.0007	-3%	0.0002	-3.54	0	-0.0011	-0.0003
geography (Census division)	0.0000	0%	0.0001	0.2	0.844	-0.0002	0.0002
size	0.0008	3%	0.0001	6.46	0	0.0005	0.0010
quintiles of occupation	0.0142	48%	0.0007	21.01	0	0.0129	0.0155
Establishment Herfindahls	0.0009	3%	0.0003	2.89	0.004	0.0003	0.0015
Establishment predicted var(ln wages)	0.0139	47%	0.0007	20.66	0	0.0126	0.0153
Unexplained (wage structure changes)							
Total	-0.0173	100%	0.0019	-8.92	0	-0.0211	-0.0135
Reweight_error	-0.0001	0%	0.0008	-0.07	0.947	-0.0017	0.0016
Pure_Unexplained	-0.0172	100%	0.0018	-9.81	0	-0.0207	-0.0138

Notes: These are the results of Rios-Avila's implementation of Fortin, Firpo, and Lemiueux's Recentered Influence Function Decomposition. Industry here is grouped into 19 supersectors and geography into 7 Census divisions. Establishment size is measured in 8 categories (1-4 employees, 5-9, 10-19, 20-49, 50-99, 100-249, 250-499, and 500+). Establishment-level Herfindahl-Hirschman indices of occupations are measured with quartiles of the distribution, interacted with occupational quintiles. The predicted variance of ln(wage) for establishments is divided into quartiles, and also interacted with occupational quintiles, with an additional dummy variable for low-wage occupations in establishment of less than 100 workers that are in the bottom half of the predicted variance distribution. Standard errors have not been bootstrapped.